Service Quality Perception Difference between Employees and Customers

NG, Wai Hung Thomas

吳偉雄

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Philosophy

in

Management

August 2002


香港中文大學
THE CHINESE UNIVERSITY OF HONG KONG
大學
統系館書圖
07 MAR 2003
LIBRARY

Abstract of thesis entitled:
Service Quality Perception Difference between Employees and Customers
Submitted by **NG Wai Hung Thomas**
For the degree of **Master of Philosophy in Management**
At The Chinese University of Hong Kong in August 2002

ABSTRACT (ENGLISH)

Studies comparing two or more parties' perceptions are abundant in the arena of social science. To practitioners, arming with a more concrete understanding of perception discrepancy (e.g., between supervisors and subordinates, male and female employees, self and peers) helps suggest where interventions can be laid or what strategies can be formulated to improve a firm's well being. One of the areas that worth much attention nowadays is the service quality (SQ) perception discrepancy existing between employees and customers. Minimizing SQ perception discrepancy between employees and customers is always paramount to satisfying customers and sustaining superior performance.

The objective of the current study is to apply measurement equivalence / invariance (ME/I) tests to examine multiple forms of SQ perception discrepancy between employees and customers. ME/I tests are to test whether measurement operation yields measures of the same attribute under different conditions. They are increasingly used to study a wide range of organizational phenomena.

It is proposed that there can be seven forms of discrepancy emerging between employees and customers which fall into the three different areas. The first area concerns with those discrepancies related to *SQ conceptualization* which is the way people conceptualize or define SQ. The two forms of discrepancy in this area are the

difference in conceptual model and *difference in manifestation of constructs*.

The second area concerns with those discrepancies pertinent to the *psychometric properties* of the measurement scale employees and customers use to evaluate the SQ. There are three particular forms of discrepancy under this area: *difference in magnitude of random error*, *difference in perception dispersion* and *difference in baseline perception*.

Finally, the third area concerns with those discrepancies germane to the *SQ evaluation*. In particular, there are two forms of discrepancy in this area: *difference in interrelationships among dimensions* as well as *difference in perceived SQ level*.

A dataset was collected from the gas station industry with an effective sample size of 350 employees and 447 customers. Using structural equations modeling (SEM), and in particular, ME/I tests, this study showed that there was no difference in the conceptual model in the two groups' SQ conceptualization. Both parties used a 2-factor structure, which consists of the dimension "extrinsic SQ" and "intrinsic SQ", as the conceptual model in their mindsets. Differences in manifestation of constructs were found for both dimensions where "extrinsic SQ" contained one non-invariant item whilst "intrinsic SQ" contained six.

Differences in magnitude of random error as well as differences in perception dispersions were found. Specifically, customers tended to have more random errors than employees in SQ evaluation, and less consensus regarding the perceived SQ level of each dimension. Also, customers showed leniency and stringency in their evaluation of SQ levels on some service components, and they displayed stringency

more often than leniency. Finally, both parties perceived no differences in interrelationships among SQ dimensions; employees however were found to have higher perceived SQ levels on both SQ dimensions, implying that they inclined to over-evaluate the SQ levels when comparing to customers. Ignorance of these multiple forms of discrepancy might eventually drive customers away.

Also as a response to the call to use latent variable structural equation modeling (LVSEM) more in consumer research by Mackenzie (2001), this paper provides a SEM approach to examine these multiple forms of discrepancy between employees and customers. This method on one hand provides neat and rich information to managers about the two parties' perceptions with a range of ME/I tests and on the other hand capitalizes the benefits of using SEM.

ABSTRACT (CHINESE)

在社會科學中，研究人員總喜愛比較兩個組別在某些事物上的看法。對於管理階層，比較兩組人的看法(例如比較上司跟下屬，男性跟女性員工，自我跟同輩等)能夠提示他們如何在公司內作出相應的策略及調整。其中值得多加注意的，是員工跟顧客在服務質素上的看法之差別。今時今日，盡量減低員工跟顧客在服務質素上的看法之差別對於滿足顧客及維持公司出色表現是非常重要的。

本研究的目的是在於應用 Structural Equations Modeling (SEM) 當中的 Measurement Equivalence/Invariance (ME/I) 測驗來研究員工跟顧客在服務質素上的看法之差別。ME/I 的測驗在研究機構中的各種情況及現象愈來愈普及。

本研究提出，員工跟顧客在服務質素上的看法之差別其實一共可分爲七種，並來自三個不同的範圍。第一個範圍是這兩組對於服務質素的概念之差別，在這個範圍之下的兩種差別分別是：一，兩組所擁有的服務質素「概念模型」之分別；二，在服務質素概念模型中，每一個層面之表現的分別。

第二個範圍是在於回應問卷問題時，員工跟客人所顯出的特徵與趨向之差別。在這個範圍之下有三種差別，分別爲：一，在回應問題時，兩組所表現出的量度錯誤之多少的不同；二，兩組對於每一個服務質素層面上之評價的共識分別；三，兩組在評審各個服務環節時，所顯出的寬容度之差別。

第三個範圍是他們在對於服務質素的評審之差別，在這個範圍之下有兩種差別，分別是：一，兩組認爲各服務質素層面的互相關係之差別；二，兩組對於每個服務質素層面上的評審之差別。

我們在汽車加油站這個服務業中，訪問了一共 335 位員工及 447 位顧客。透過 ME/I 的測驗，我們發現員工跟顧客在服務質素的概念模型上沒有分別。顧客及員工皆採用一個包含兩個服務質素層面的概念模型，而這兩個層面分別被稱爲「內在」及「外在」服務質素。另外，在這兩個層面的表現上，兩組皆認爲除了七個服務環節外，其他服務環節與其代表的服務質素層面的關係都是一樣的。

跟員工相比，顧客在回應問卷問題時，傾向顯示出較多的量度錯誤，與及在剛才提到的兩個服務質素層面上擁有較低的共識。另外，在某些服務環節的評審上，顧客跟員工有不同的寬容量。最後在服務質素的評審上，顧客跟員工對於「內在」及「外在」服務質素這兩個層面的關係是沒有分歧的，但對於這兩個層面上的服務質素之評審，顧客跟員工則有顯著的差別。假若管理層漠視這種種差別，顧客或許會認爲所提供的服務不能完全滿足他們的要求而離開。

本文章提出利用 SEM 及 ME/I 的測驗來分析存在於顧客及員工之間，對服務質素上的看法之差別。這方法一方面能夠幫助管理人員收集整齊及豐富的資料；另一方面亦能套現 SEM 所帶來的好處。本研究亦是回應 MacKenzie (2001) 的催促，要多用 SEM 於市場學的研究上。

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Dr. Gordon Cheung for his invaluable guidance in preparing this thesis over the last two years. He also showed his wholehearted care in helping me develop to be a potential researcher. I am more than thankful that he was willing to be my supervisor.

Special thanks also have to be given to Dr. Michael Hui who also provided me with a range of invaluable opinions and suggestions in improving the quality of this thesis. I enjoyed so much in working with these two outstanding scholars.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURE

CHAPTER I

INTRODUCTION

Different people see things in different ways. Perception discrepancy across groups is indisputably always of enormous interest to researchers. Studies comparing two or more parties' perceptions are abundant in the arena of social science. Researchers might contrast the perceptions between male and female (e.g., Franke, Crown & Spake, 1997), supervisors and subordinates (e.g., Johnson, 2000), across races (e.g., Chan, 1997), nationalities (e.g., Nishida, Hammer & Wiseman, 1998), cultures (e.g., Fok, Hartman, Villere & Freibert, 1996) or so forth. Examining perception discrepancy is paramount because it could augment our understanding of world phenomena.

In a practical perspective, perception discrepancy suggests to managers where interventions can be laid or what strategies can be formulated to improve a firm's well being. For instance, subordinates are found to perceive significantly lower levels of room for creativity on their jobs when comparing to supervisors (Johnson, 2000). Acknowledging the existence of this type of perception discrepancy, managers and supervisors might consider permitting more opportunities for subordinates' creativity on their jobs so as to enhance their job satisfaction. Since understanding perception discrepancy is useful, researchers and practitioners should emanate as much information as possible from comparing two (or more) parties' perceptions. This paper aims at suggesting that perception discrepancy can indeed be examined in a more detailed manner. Specifically, it is proposed that perception discrepancy can

take many forms. Each form of discrepancy could convey to managers a piece of unique information about the two parties' perceptions.

One of the areas that would benefit from examining several forms of perception discrepancy is the service quality (SQ) perception discrepancy existing between employees and customers. SQ is the ability of an organization to meet customers' expectations (Parasuraman, Zeithaml & Berry, 1985, 1988). Every organization *has* to pay attention to SQ nowadays in order to stay survive in the market, let alone outperform rivals (Mattsson, 1994). Managers have to be alert that keeping satisfying SQ is a sine qua non for competitive business performance (McDougall & Levesque, 1994). Ensuing the pioneering work of Parasuraman et al. (1985, 1988), SQ at soon became an intensive area of research. There were over 4000 articles in the last two decades that mentioned "service quality" (Philip & Hazlett, 1997).

In the SQ literature, one of the sparsely examined areas is SQ perception discrepancy existing between employees and customers (Bitner, Booms & Mohr, 1994). However, this research area should not be ignored as curtailing any SQ perception discrepancy is paramount to providing excellent quality service (Klose & Finkle, 1995; McColl-Kennedy & White, 1997). Managers and employees might think the SQ is fine; but if customers disagree, the company has a problem (Parasuraman & Berry, 1991). The emergence of SQ perception discrepancy between employees and customers might eventually drive the customers away. Thus, reduction of SQ perception discrepancy existing between these two groups is necessary for a firm's superior performance nowadays.

In this paper, it is postulated that SQ perception discrepancies existing between employees and customers indeed fall into three areas. The first area is those discrepancies related to the *SQ conceptualization* which is the way people conceptualize the SQ provided by an organization (e.g., Parasuraman et al. 1988). SQ conceptualization discrepancy emerges when employees and customers do not agree on the way they conceptualize or think of SQ. The second area is pertinent to those discrepancies the two parties display on the psychometric properties of the scale they use to evaluate SQ. Inasmuch as the two groups might not respond to the same measurement scale in the same way, discrepancies in this area might be observed.

The third area is those discrepancies germane to *SQ evaluation*. One of the most common forms of discrepancy emerged in this area that interest researchers is the difference between employees and customers in the *perceived SQ level* of the service a firm provides. Perceived SQ level, in the simplest terms, is the judgment about a service firm's excellence or superiority (Zeithaml, 1987). It can be described as "high" (vs. "low") or "good" (vs. "bad"). These three areas represent three different facets of SQ perception discrepancy. SQ conceptualization discrepancies focus on the differences between employees and customers in how they *think of* or *conceptualize* the SQ; discrepancies in psychometric properties of the measurement scale focus on the differences between employees and customers in how they *respond* to the SQ measurement scale; SQ evaluation discrepancies focus on the differences between the two groups in how they *evaluate* and *judge* the SQ.

In the body of literature, only a few of studies have given attention to examine SQ perception discrepancies between employees and customers. These researchers

mainly concern with the difference in the perceived SQ level between the two groups (e.g., Klose & Finkle, 1995; Reynierse & Harker, 1992; Schneider & Bowen, 1985). On the other hand, almost no studies have examined the discrepancies in the *SQ conceptualization* and *psychometric properties* of the measurement scale existing between employees and customers. Thus, the objective of this study is to stimulate the research in comparing employees' and customers' SQ conceptualization, psychometric properties of the measurement scale and SQ evaluation by proposing several forms of discrepancy between employees and customers emerging from these three areas.

The several forms of discrepancy proposed in this paper correspond to several measurement equivalence/invariance (ME/I) tests in structural equations modeling (SEM) application. ME/I is "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn & McArdle, 1992). Applying it into the context of the current study, it is a question of whether or not employees and customers show equivalence of measures in the SQ conceptualization, psychometric properties of the measurement scale and SQ evaluation. ME/I tests are increasingly used to examine organizational phenomena in a variety of contexts (Vandenberg & Lance, 2000).

With the acknowledgement of the multiple forms of discrepancy in SQ conceptualization, psychometric properties of the measurement scale and SQ evaluation existing between employees and customers, managers will be armed with a more concrete understanding of the particular differences existing between the two groups. These multiple forms of discrepancy provide with managers a range of

information that could help them reconcile employees' and customers' perspective, and improve the service in the competitive service market. Ignorance of these discrepancies emerging from the aforementioned areas might deteriorate a firm's business performance.

For instance, SQ conceptualization suggests to managers the assessment criteria of SQ (Brady & Cronin, 2001). If SQ conceptualization discrepancies between employees and customers are unnoticed, customers might regard the employees or the firm cannot fully fulfill their criteria and expectations, and might judge the delivered SQ as poor. When they are not completely satisfied, the firms might be in jeopardy of losing the customers (Klose & Finkle, 1995). And this might even trigger off other undesirable consequences like bad word-of-mouth and reduced profitability. Therefore the repercussion of ignoring the SQ conceptualization discrepancies emerging between employees and customers can be serious. Similarly, attention must also be given to the possible forms of discrepancy existing in the other two areas mentioned.

CHAPTER II

LITERATURE REVIEW

Service Quality (SQ)

SQ is the ability of an organization to meet customers' expectations (Parasuraman et al., 1985, 1988). The construct "service quality" and SERVQUAL (Parasuraman et al., 1988), which is the most commonly used scale to measure SQ, had stormed the marketing literature in the last two decades (Johnston, 1995). Perhaps it is because the concept "service quality" is in nature elusive and difficult to grasp (Brady & Cronin, 2001) as it is of three unique inherent characteristics: intangibility, heterogeneity and inseparability (Parasuraman et al., 1985). A review of literature shows that research pertaining to SQ usually revolve around at least four major issues.

The first area is those research that try to identify the determinants (e.g., Mersha & Adlakha, 1992; Parasuraman et al., 1985) and consequences of SQ (e.g., Grapentine, 1998; Zeithaml, 2000). For instance, a strong service climate instilled in an organization is found to lead to a better overall SQ evaluation from customers (Schneider, White & Paul, 1998). Positive SQ evaluation from customers is found to lead to a higher willingness to recommend the company, purchase intention, re-purchase, loyalty, and ultimately a firm's profitability (Iacobucci, Grayson & Ostrom, 1994; Parasuraman et al., 1988; Zeithaml, Berry & Parasuraman, 1996).

The second area of research interest revolving SQ roots in its seemingly similar nature with customer satisfaction. It therefore spurs a range of research to examine the differences between the two constructs (e.g., Bolton & Drew, 1991; Iacobucci, Ostrom & Grayson, 1995; Zeithaml, Berry & Parasuraman, 1993) and the direction of causality between the two (e.g., Lee, Lee & Yoo, 2000). Regarding the differences between the two constructs, there appears to be a consensus up to date that while satisfaction refers to the outcome of the service transaction, SQ represents the customer's impression of the relative superiority (or inferiority) of the organization and its service (Johnston, 1995). In other words, satisfaction describes a more specific and short-term evaluation and SQ describes a more general and long-term evaluation (Bitner & Hubbert, 1994). Following that, customers satisfaction should then be the antecedent rather than the consequence of SQ as repeated satisfaction of specific service encounters logically leads to a general and long-term SQ evaluation (Parasuraman et al., 1988). However, other researchers suggest it should be the other way around (e.g., Dabholkar, et al., 2000; Spreng & Mackoy, 1996) as SQ could be the antecedents to customer satisfaction regardless of whether these constructs were measured for a given experience or overtime (Oliver, 1993). It seems that more research have to be done before an affirmative conclusion can be obtained.

The third area is pertinent to the controversial SERVQUAL scale. As attracting attention as the construct itself, the SERVQUAL scale had become one of the most debated issues in the SQ literature. This model suggests that when people think of SQ, they usually embody five dimensions: tangibles, reliability, responsiveness, assurance and empathy (Parasuraman et al., 1988). Regardless of the industry-type, reliability is usually the most important dimension to customers, followed by

responsiveness, assurance and empathy. The tangibles are usually of the least concern to customers (Zeithaml, Parasuraman & Berry, 1990). Table 1 provides a brief explanation of what each dimension particularly refers to.

TABLE 1

Definitions of SERVQUAL Dimensions

(Parasuraman, Zeithaml, & Berry, 1988)

Dimension	Definition
Tangibles	Physical facilities, equipment, and appearance of personnel
Reliability	Ability to perform the promised service dependably and accurately
Responsiveness	Willingness to help customers and provide prompt service
Assurance	Knowledge and courtesy of employees and their ability to inspire trust and confidence
Empathy	Caring, individualized attention the firm provides to its customers

The 22-item SERVQUAL scale is operated by using difference scores as the authors of the scale argue that perception of SQ comes from a comparison of expectations and perceived performance (Parasuraman et al., 1988). Thus, in administering SERVQUAL scale, each item is used twice: first, to measure customers' expectations in general; second, to measure perceptions of performance of a particular firm. An overall SQ index score for a dimension can be obtained by averaging all those items representing that dimension.

Researchers keep posing challenges to the scale about its replicability in other industries as researchers are seldom able to generate the same SERVQUAL structure in other industries of interest (e.g., Cronin & Taylor, 1992a; Durvasula, Lysonski &

Mehta, 1999). Also, the five dimensions of SERVQUAL are rather unstable and therefore its dimensionality has also been questioned repeatedly (e.g., Mels, Boshoff & Nel, 1997; Robinson, 1999). In addition, there are controversies regarding its basis of measurement. While the authors of the SERVQUAL vociferate that the scale should be operated by performance-minus-expectation scores, others suggest the usage of only performance-score (e.g., SERVPERF; Cronin & Taylor, 1992b); the incorporation of importance score (Carman, 1990) and other modifications of the scale (e.g., Teas, 1993).

Finally, the last area that have aroused SQ researchers' interest, yet is the less frequently examined area (Bitner et al., 1994) when comparing to others, is the comparison of SQ perceptions between customers and service providers. In particular, these researchers are interested in comparing the perceived SQ levels between the two groups (e.g., Brown & Swartz, 1989; Schneider & Bowen, 1985). On the other hand, almost no studies have attempted to examine the discrepancies between employees and customers in the *SQ conceptualization* and the *psychometric properties* of the measurement scale they use to evaluate SQ. As mentioned before, examining the multiple forms of discrepancy existing between employees and customers in these three areas is important to a firm's well being. In spite of that, empirical research in this arena are still sparse.

Conceptualizing SQ

How people conceptualize SQ has attracted researchers' attention since the business market has become mainly service-based. However, there is not much

convergence towards which particular forms of model could appropriately represent SQ conceptualization, as SQ is intangible and therefore elusive (Carman, 1990). Most researchers and practitioners use the 22-item SERVQUAL scale as a manifestation of how people conceptualize SQ. As stated, this model embodies five SQ dimensions: tangibles, reliability, responsiveness, assurance and empathy (see table 1). Although the author validated the SERVQUAL scale in a range of industries, the scale was still "unstable" in its structure as researchers incessantly found SQ models at variation with SERVQUAL structure. Table 2 provides a review of those studies that had attempted to replicate the factor structure of SERVQUAL.

TABLE 2

SERVQUAL Replication Studies

Studies	Survey Instrument	Factor Structure
Babakus & Boller, 1992	Original 22 items	5-factor was not supported. 2-factor solution (but determined by the directions of item wording)
Bresinger & Lambert, 1990	Original 22 items	19 items resemble SERVQUAL's 5 dimensions, though only 4 have eigenvalues > 1
Carman, 1990	12-21 items (in 4 industries)	5- to 9-factor structure
Caruana et al., 2000	Revised 21 items (Parasuraman et al., 1994)	3-factor: (1)tangibles, (2)reliability, (3)responsiveness, assurance and empathy load on the third factor
Cronin & Taylor, 1992a	Original 22 items	Unidimensional
Durvasula et al., 1999	Original 22 items	3-factor, though the fit is marginal: (1)tangibles, (2)reliability, (3)responsiveness, assurance and empathy load on the third factor

TABLE 2 (Continued)
SERVQUAL Replication Studies

Studies	Survey Instrument	Factor Structure
Finn & Lamb, 1991	Original 22 items	5-factor model has a poor fit
Gagliano & Hathcote, 1994	Original 22 items	4-factor: Personal attention, reliability, tangibles, convenience
Lam, 1997	Original 22 items	Unidimensional
Mels et al., 1997	Original 22 items	2-factor: Intrinsic SQ, Extrinsic SQ
McDougall & Levesque, 1994	Original 22 items	3-factor: Tangibles, contractual performance, customer-employee relations
Parasuraman et al, 1991	Original 22 items	5-factor, but tangibles split into 2 factors while responsiveness and assurance load on 1 factor
Schneider et al., 1992	Original 22 items	3-factor: dependability/trust, personal attention/helpfulness, equipment/facilities
Spreng & Singh, 1993	Original 22 items	5-factor model has a poor fit
Yavas, 1998	Original 22 items	2- to 4-factor (depending on industry)

Among them, it seems that only Brensinger and Lambert (1990) were able to replicate a factor structure *closest* to that of SERVQUAL. They found that with the exception of 3 items, the remaining 19 items loaded on five factors that represent the SQ dimensions postulated in SERVQUAL, though only four factors had eigenvalues larger than one. Most of the rest of the studies however failed to replicate the SERVQUAL's factor structure. Specifically, the factor structure of the SERVQUAL was found to vary from one (Cronin & Taylor, 1992a; Lam, 1997), two (Babakus & Boller, 1992; Mels et al., 1997), three (Caruana, Weing & Ramaseshan, 2000; McDougall & Levesque, 1994; Parasuraman, Zeithaml & Berry, 1994; Schneider,

Wheeler & Cox, 1992;), four (Gagliano & Hathcote, 1994; Yavas, 1998), five (Bresinger & Lambert, 1990; Parasuraman, Berry & Zeithaml, 1991) and more than five factors (Carman, 1990).

Increasing evidences show that special attention should be given to the "things" underlying the terms described by SERVQUAL dimensions (Brady & Cronin, 2001). In other words, the five dimensions of SERVQUAL are terms that might be used to describe some important aspects of SQ. However, of even greater concern is indeed "what" should be reliable, responsive, empathic, assured and tangible (Brady & Cronin, 2001). This gives rise to another stream of research which postulates that SQ conceptualization is better represented by Grönroos's (1984) concept that SQ should consist of mainly two dimensions, the SQ pertinent to the *process* and the SQ pertinent to the *outcome* (e.g., McDougall & Levesque, 1994). Specifically, Grönroos (1984) suggest the following two SQ dimensions: functional quality and technical quality. Functional quality is people's perceptions of the interactions that take place during service delivery. Technical quality on the other hand reflects what the customer received in the service encounters. It seems that the outcome aspect of SQ is missing in SERVQUAL dimensions (Bienstock et al., 1997).

In light of this direction, researchers try to propose models that incorporate the concepts of *process* and *outcome* SQ. For example, Rust and Oliver (1994) suggested a SQ model that embodied three dimensions: service product (the technical quality), service delivery (the functional quality) and service environment. Similarly, McDougall and Levesque (1992) found three dimensions for SQ conceptualization in their study: tangibles, contractual performance and customer-employee relations.

They concluded that perceived SQ had two main facets, one representing the output quality (contractual performance in their study), the other the service process (customer-employee relations in their study). Mels et al. (1997) similarly proposed that SQ should consist of extrinsic SQ and intrinsic SQ which to some extent resemble the technical and functional quality respectively. Other studies have also suggested SQ conceptualization models that place emphasis on both the quality of *process* and *outcome* (e.g., Bienstock, et al., 1997; Genestre & Herbig, 1996).

Inspired by Grönroos's (1984) idea and in the face of the mixed factor structures of SERVQUAL, Brady and Cronin (2001) believed that a multi-level model might be able to solve these quandaries about SQ conceptualization as SQ evaluations are highly complex and might operate at several levels of abstraction (Carman, 1990). They therefore proposed that "service quality" should consist of three sub-dimensions: interaction quality, outcome quality and physical environment quality, with each of these three sub-dimensions possesses three finer sub-dimensions too.

Regardless of the stream researchers and practitioners choose to represent how people conceptualize SQ, it seems reasonable to conclude that first, SQ conceptualization is still of paramount interest to researchers as understanding how people, especially customers, conceptualize SQ is like understanding how they assess the quality of the service (Brady & Cronin, 2001). Second, there is still no consensus regarding which SQ conceptualization models could best represent how people conceptualize SQ.

SQ Perception Discrepancy between Employees and Customers

SQ Conceptualization Discrepancy. In the SQ literature, scant attention had been given to compare employees' and customers' *SQ conceptualization* (see the previous section for a review of the studies that have examined how people conceptualize SQ). Inasmuch as employees and customers might not define or conceptualize SQ in the same way, discrepancies in this area might be observed between the two groups.

In review, no studies have purportedly compared employees' and customers' SQ conceptualization. Only Bitner, Booms and Mohr (1994) as well as Genestre and Herbig (1996) have *indirectly* thrown light on this subject. Bitner et al. (1994) compared the satisfying and dissatisfying incidents reported by employees and customers. In other words, they contrasted how the two groups would *think* about the service encounters. Using critical incidents technique which is a systematic procedure for recording events and behaviors that are observed to lead to success or failure on a specific task (Ronan & Latham, 1974), they found that employees and customers recounted similar categories of satisfying and dissatisfying incidents in three service industries (restaurants, airlines and hotels). This result hints that employees and customers might share a rather similar SQ mindset or SQ conceptualization.

Genestre and Herbig (1996) also indirectly shed light on the equality of SQ conceptualization between employees and customers, though this issue was not one of the objectives of their study. In proposing a SQ model different from SERVQUAL,

they collected data from employees and customers in order to validate the model. When they presented the result of the factor structure obtained from employees and customers data, differences were found. The authors did not elaborate on it as the comparison between employees and customers in the SQ conceptualization was not of importance to their study. However, it is important to the current study as it evidences that the two groups might not always have the same SQ conceptualization otherwise.

Discrepancy in Psychometric Properties of SQ scale. In the SQ literature, there are apparently no studies that have examined the differences in the psychometric properties of the measurement scale employees and customers use to evaluate SQ. Thus, little is known so far on how the two groups would differ in the way they respond to the same SQ measurement scale.

SQ Evaluation Discrepancy. While little effort has been paid to compare employees' and customers' SQ conceptualization and psychometric properties of the measurement scale in the literature, more studies have placed effort in examining whether the two groups possess a similar SQ evaluation, and in particular *perceived SQ level* which is the judgment about a service firm's excellence or superiority (Zeithaml, 1987). Yet, the answer to this research question is still far from conclusive. Table 3 provides a summary of those studies that have compared the two parties' perceived SQ levels.

TABLE 3

Past Studies on Comparing the Perceived SQ Levels between Employees and Customers

Article	Context	Methodology	Findings
The following studies found that employees' and customers' perceived SQ levels were similar			
Schneider et al., 1980	Examining perceived SQ level of bank service	Correlation	Customers' and bankers' perceived SQ level are positively related
Schneider & Bowen, 1985	Examining perceived SQ level of bank service	Correlation	Customers' and bankers' perceived SQ level are positively related
Reynierse & Harker, 1992	Examining perceived SQ level of bank service	Correlation	Customers' and bankers' perceived SQ level are positively related
Klose & Finkle, 1995	Examining the relationship between service-provider gap and service quality gap	Correlation	Employees' and customers' perceived SQ levels are positively related

TABLE 3 (Continued)

Past Studies on Comparing the Perceived SQ Levels between Employees and Customers

Article	Context	Methodology	Findings
The following studies found differences in the perceived SQ levels between employees and customers			
Brown & Swatz, 1989	Examining perceived SQ level of professionals and their clients	Comparing the mean scores of each pair of same statements	Inconsistency exists between two parties' perception
Comila, 2000	Comparing perceived SQ level between customers and employees in India	Comparing the mean scores of each pair of same statements	There a lot of differences between the two groups
Hopkins et al., 1993	Comparing perceived SQ level between shippers (customers) and carriers (employees)	Comparing the mean scores of each pair of same statements	There is a significant difference between two parties' perceived SQ level
McColl-Kennedy & White, 1997	Comparing perceived SQ level of hotels between customers and employees	Comparing the mean scores of each pair of same statements	Mismatch of perceived SQ level between customers and employees

Researchers tried to pore over the relationship between employees' and customers' perceived SQ levels since decades ago. Schneider, Parkington and Buxton (1980) pioneered to show that employees' perceived SQ level had a significant positive correlation of 0.67 with that of customers. Along the same line, other researchers replicated this finding and concluded that employees' perceived SQ level, or attitude towards SQ, was positively correlated to customers' (Klose & Finkle, 1995; Reynierse & Harker, 1992; Schneider & Bowen, 1985). These studies render evidences that employees and customers might usually perceive the quality of a service rather similarly: when employees think that the SQ they have a part to provide is good, customers are likely to agree so.

However, this hardly is an affirmative conclusion (Bitner et al., 1994). Other studies on the contrary suggest an opposite inference. Brown and Swartz (1989) compared patients' service experiences to the physicians' perceptions of their patients' experiences. They found that inconsistency existed between the two parties' perceived SQ levels. In the same vein, it was found that carriers (the employees) did not have the same perceived SQ level as shippers (the customers) (Hopkins, Strasser, Hopkins & Foster, 1993); the perceived SQ level judged by hotel employees was found to be different from that judged by hotel customers (McColl-Kennedy & White, 1997); there were also disagreements between employees' and customers' perceived SQ level in foreign exchange and corporate travel industry in India (Comila, 2000). These, contradictory to the aforementioned findings, demonstrate that employees might not always have the same perceived SQ level as customers do. This gap imperils a firm's profitability and competence.

The antithetical conclusion from the literature might suggest that the equality of perceived SQ levels between employees and customers is contingent on a lot of determinant factors. One of them is the type of industry. For example, in the service industries where there are no strong "role scripts" (Soloman, Suprenant, Czepiel & Gutman, 1985) for both employees and customers, perceived SQ levels might be very different between the two groups; for industries where the role scripts are well-defined, the SQ level *could* not go too far away from the average standard or expected level. Thus, the perceived SQ level should not show too much of differences between the two groups (Bitner et al., 1994).

From the above, two observations could be made regarding the past literature in examining the discrepancies between employees and customers in the three mentioned areas. First, only a few of studies have examined the discrepancies existing between employees and customers in the area of *SQ evaluation*. Among them, most of the effort has been placed on investigating the equality of the perceived SQ levels between the two parties (see table 3). However, mixed conclusions were found on this research question. Second, almost no studies have attempted to examine discrepancies emerged from areas other than *SQ evaluation*. As narrated, only Bitner et al. (1994) and Genestre and Herbig (1996) have *indirectly* examined the differences in the SQ conceptualization between employees and customers. The differences in the psychometric properties of the measurement scale between the two parties have been seemingly ignored.

Measurement Equivalence / Invariance (ME/I)

As articulated before, measurement equivalence/invariance (ME/I) tests will be applied to examine the several forms of discrepancy proposed in this study. ME/I is "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn & McArdle, 1992). Structural equations modeling (SEM) and in particular ME/I tests are increasingly adopted in a gamut of areas. Examples of these are assessment of longitudinal change (e.g., Vandenberg & Self, 1993), cross-cultural comparisons (e.g., Cheung & Rensvold, 2000; Steenkamp & Baumgartner, 1998), and self-others ratings comparison in organizations (e.g., Cheung, 1999).

When examining ME/I between two groups, eight ME/I tests in total can be performed (Vandenberg & Lance, 2000). However, the first test, which examines the equivalence of covariance matrices between two groups, is usually not executed as this test is expected to "fail" because it is difficult to get two equal covariance matrices from two different groups. On the other hand, the seven tests explicated below are performed more often in organizational research. Vandenberg and Lance (2000) provide a detailed review on the ME/I tests literature. The following description of these seven tests is mainly based on their review.

Testing Configural Invariance. This is to test the null hypothesis that two parties (e.g., employees and customers in this study) share the same pattern of fixed and free factor loadings (Horn & McArdle, 1992). In other words, it is to test whether two groups have an equivalent factor structure. Because factor structure is a reasonable empirical map of the underlying cognitive mindset, a difference in factor structure signalizes a significant difference between two groups (Vandenberg &

Lance, 2000). Therefore, this null hypothesis must not be rejected for further tests to go on as it is meaningless to compare other properties of the factor structure when the two groups are not referring to the same underlying constructs (Vandenberg & Lance, 2000).

Testing Factorial Invariance. This is to test the null hypothesis that factor loadings for like items are invariant between two groups (i.e., $\Lambda^{(1)} = \Lambda^{(2)}$, where the parenthetical superscript represents the group membership; Horn & McArdle, 1992). It is the most frequently performed test in a review of ME/I literature (Vandenberg & Lance, 2000). Factor loadings are the regression slopes relating the items to the corresponding underlying latent variables. They are thought to reflect the relative strength of relationships between the indicators and the underlying latent variables (Cheung, 1999). A factor loading represents the expected change in the score of the observed item with per unit change in the respective underlying construct.

Adding the constraint of factorial invariance is a stronger (more restrictive) test than the configural invariance test alone in that the equality constraint is further imposed on the values of all factor loadings in addition to the factor structure. However, since it is hard to obtain full factorial invariance for all items between two groups, partial factorial invariance (Byrne, Shavelson & Muthén, 1989) always has to be invoked (The number of non-invariant items should not be too large, or it is hard to argue that both groups are referring to the same frame of reference; Cheung & Rensvold, 2000). This essentially means allowing non-invariant items to stay "free" in the model after they were identified. Configural invariance and at least partial factorial invariance must be established for further tests to go on or again it is

difficult to argue that the two parties are referring to the same frame of reference (Cheung, 1999).

Testing Unique Variance Equivalence. This is to test the null hypothesis that like items' unique variances are equivalent between two groups ($\Theta^{(1)} = \Theta^{(2)}$). SEM allows the estimation of the unique variance for each item which is the non-systematic error incurred to respondents in the measurement process. A difference in unique variance (on a particular item) shows that one of the groups tends to exhibit a larger magnitude of random error in their evaluation scores (given on that particular item).

Testing Factor Variance Equivalence. This is to test the null hypothesis that factor variances are invariant between two groups ($\Phi_j^{(1)} = \Phi_j^{(2)}$ where the subscript j corresponds to *jth* factor). Factor variance represents the dispersion of a latent variable. Thus, rejection of this null hypothesis signalizes that the group with the larger factor variance uses a wider range of the construct continuum to respond to the indicators (Vandenberg & Lance, 2000). Therefore, this test can be used to test the equality of the degree of consensus in the evaluation scores given by two groups.

Testing Intercept/Scalar Invariance. This is to test the null hypothesis that intercepts of like items' regression on the underlying latent variable are invariant between two groups (i.e., $\tau^{(1)} = \tau^{(2)}$). It is the least frequently examined test in the ME/I tests literature (Vandenberg & Lance, 2000). Because an intercept in nature represents the value of an observed item when the value of the underlying latent

variable is zero, this test has been used to test for the existence of systematic response style, for instance, leniency or stringency (e.g., Cheung & Rensvold, 2000). On the other hand, differences in the intercept terms might also reflect two groups' differences in the level of the construct (Vandenberg & Lance, 2000). If that is the case, significant differences between two groups in all items' intercept terms (for a dimension) should be expected.

Testing Factor Correlation Equivalence. When testing the null hypothesis that factor covariances are invariant between two groups ($\Phi_{ij}^{(1)} = \Phi_{ij}^{(2)}$ where i and j represent two different factors, i.e., latent variables), one must be aware that factor covariance is a function of factor variance. If factor variance equivalence is rejected for a factor, it is reasonable to expect factor covariances between that particular factor and others to be different between the two groups (Cheung, 1999; Vandenberg & Lance, 2000). Therefore, researchers usually test for the factor correlation equivalence instead. Testing the equality of factor correlations in essence examines whether two groups perceive the same magnitude of inter-relationships among factors.

Testing Latent Means Equivalence. Finally, this test is to examine the null hypothesis that the means (latent means) of like factors are equal between two groups (i.e., $\kappa^{(1)} = \kappa^{(2)}$). This test is similar to the traditional ANOVA or t-tests where the interest is in examining if the two groups have shown substantive differences in the level of the variable (Vandenberg & Lance, 2000).

The above seven ME/I tests are commonly executed. Indeed, as Cheung (1999)

implies, every one of these tests has its own "conceptual meaning" that could convey to researchers a piece of unique information about the two groups. This study also tries to apply the seven ME/I tests to examine several forms of discrepancy existing between employees and customers in the area of SQ conceptualization, psychometric properties of the measurement scale and SQ evaluation. The "conceptual meaning" of each form of discrepancy could help managers reconcile the two perspectives.

CHATPER III

OBJECTIVE

How to incessantly improve SQ is a constant challenge to managers. The employees-customers interface still remains the most significant determinant of customers' SQ judgment (Hartline & Ferrell, 1996). Consumers are reluctant to complain when they are dissatisfied with the service (Brown & Swartz, 1989). The existence of any form of discrepancy between employees and customers in SQ conceptualization, psychometric properties of the measurement scale and SQ evaluation may not be noticed until it is too late to implement remedy (McAlexander, Kaldenbugy & Koenig, 1994).

From the literature review, it is however found that only few studies have attempted to examine discrepancies between employees and customers in the mentioned three areas. Among these studies, most effort has been directed to examine the differences in the perceived SQ levels between employees and customers only, despite there can actually be discrepancies emerging from other areas (e.g., SQ conceptualization and psychometric properties of the measurement scale). Therefore, the objective of this study is to **advance the field by applying ME/I tests to examining multiple forms of discrepancy existing between employees and customers in the area of SQ conceptualization, psychometric properties of the measurement scale they use to evaluate SQ and SQ evaluation**. What each area particular embraces as well as the several forms of discrepancy included in these areas are explicated in the section of "Conceptualization".

These multiple forms of discrepancy allow managers to get more concrete information about the two parties in the three areas mentioned. If managers can utilize these pieces of information and achieve a close match between the two perspectives, they are more likely to provide outstanding SQ (Klose & Finkle, 1995; McColl-Kennedy & White, 1997), survive and outperform their rivals in this fiercely competitive business market. On the other hand, neglecting the possible existence of these forms of discrepancy might lead to poor customer satisfaction and reduced customer base. For instance, neglecting the difference in SQ conceptualization is like neglecting the fact that employees do not understand customers' assessment criteria of SQ. Consequently, customers might choose to switch to other rivals. Thus, managers should endeavor to keep these possible discrepancies between employees and customers minimal in order to circumvent the potentially high costs of mismatched perceptions, including losing the customers.

CHAPTER IV

CONCEPTUALIZATION

In this section, multiple forms of discrepancy that can exist between employees and customers will be introduced. In particular, it is postulated that discrepancies between the two groups can take seven forms, corresponding to seven ME/I tests mentioned previously (see Literature Review). Each form of discrepancy renders unique and useful information to managers. As aforementioned, these discrepancies can indeed fall into three areas: *SQ conceptualization*, *psychometric properties* of the measurement scale, and *SQ evaluation*.

SQ Conceptualization

SQ conceptualization, as reviewed, concerns with how people think or define the SQ a firm provides. Being able to fathom both employees' and customers' SQ conceptualization is important. If employees are found to have a different SQ conceptualization from that of customers, the company might have a problem. It signalizes that employees and customers embrace quite distinct conceptions to represent SQ. It can be surmised that they have different criteria or dimensions to assess the SQ. If managers do not grapple with this, it is likely that customers would regard the firm unable to meet their expectations, and might leave eventually. There are two forms of SQ conceptualization discrepancy: *difference in conceptual model* and *difference in manifestation of constructs*.

Difference in Conceptual Model. This form of discrepancy corresponds to

examining configural invariance (i.e. factor form invariance) in the ME/I literature where the interest is in investigating whether both groups share the same conceptual model (factor structure).

It is believed that in people's mindsets, there is generally a particular "model" of how they define SQ. Numerous studies have attempted to document the SQ models in people's mindsets (see "Conceptualizing SQ" in the literature review section). The most common model researchers and practitioners would adopt to represent how people might conceptualize SQ is the SERVQUAL (Parasuraman et al., 1988) which suggests that there are five SQ dimensions on which people could evaluate (see table 1). Managers should be relieved if employees have the same "SQ model" as customers do as it indicates that employees in general understand how customers assess the SQ.

There is a difference in SQ conceptual model when employees and customers embrace a different number of SQ dimensions in their mindsets, or when they adopt a different pattern of items to represent these SQ dimensions (Cheung, 1999; Vandenberg & Lance, 2000). In other words, the conceptual model employees and customers use to represent SQ "look" differently. For example, when thinking of SQ, customers might incorporate only one overall dimension (i.e., "service quality") in their mindset (Cronin & Taylor, 1992a), whereas employees might embrace five SERVQUAL dimensions (Parasuraman et al., 1988). Or when customers think that a particular item is representing a dimension called "customer-oriented", employees might regard that item is representing a dimension called "product confidence" (Genestre & Herbig, 1996).

There can be a lot of reasons inducing a difference in SQ conceptual models between employees and customers. Possible causes might include the influence of customers' personal service philosophies which are the "underlying generic attitude towards the meaning of service and the proper conduct of service providers" (Zeithaml et al., 1993). For instance, customers might conceptualize the SQ with two criteria, quality of those things they can see and those they cannot see. Therefore, in their conception, SQ simply consists of only two dimensions, namely tangibles and intangibles whereas employees are taught and trained to focus on the intangible aspects and therefore are more prone to see more dimensions pertaining to intangibles (e.g., reliability, responsiveness, assurance and empathy in SERVQUAL).

Experience might also affect SQ conceptualization (Cadotte, Woodruff, & Jenkins, 1987). For instance, customers might base their SQ conceptualization of a particular service firm on their experience in all similar firms whereas employees might mainly base their SQ conceptualization on their experience with the firm they are working in, engendering in different SQ models in their mindsets. Other possible reasons generating distinct SQ conceptual models between employees and customers might include customers' lack of knowledge on the service; discrepancy between the firm's external communication and service delivered (Zeithaml et al., 1990); different best-brand (Cadotte et al., 1987) taken by employees and customers as the referent in their perceptions. The presence of a difference in SQ conceptual model between the two groups precludes the examination of other forms of discrepancy narrated below as the two groups are referring to two distinct conceptions that are no longer comparable.

Difference in Manifestation of Constructs (Dimensions). Examining this form of discrepancy is tantamount to examining factorial invariance in the ME/I literature. In other words, the interest is in knowing whether two groups (e.g., employees and customers) perceive the same relative strength of relationship between the items and the underlying dimensions (in this study, the SQ dimensions).

When managers find that employees and customers share the same SQ conceptual model, they should then examine whether the two parties have differences in their manifestation of the SQ dimensions. A dimension, or a construct, is manifested by items. For instance, one of the items representing the dimension "tangibles" in SERVQUAL is "XYZ firm has modern-looking equipment" (Parasuraman et al., 1988).

Even though employees and customers might have no difference in conceptual model, they might not agree on how each construct should be manifested. Specifically, employees and customers might not perceive the same relative strength of relationship between the behaviors / events / roles (mentioned in the survey items) and the respective SQ dimensions. For example, both employees and customers might have an equal conceptual model, say SERVQUAL, in their mindsets, but customers might think that for the dimension "reliability", the item "when customers have problems, the firm should be sympathetic and reassuring" should have the largest strength of relationship with this dimension whereas employees might think that "keeping records accurately" should be the one with the largest strength of relationship. This form of discrepancy tells managers which particular facets of

service are perceived to be more closely related to the underlying dimensions by employees and customers.

Empirical evidences bolster that there is such a difference between employees and customer in the perceived relative strength of relationships between service attributes and the underlying dimensions. For example, they disagree on which attributes of hotel service are the most paramount to overall SQ (McColl-Kennedy & White, 1997). Usually the reasons governing why employees and customers have differences in manifestation of constructs should be similar to the reasons inducing a difference in conceptual model (Cheung, 1999) as both forms of discrepancy concern with the frame of reference the two groups deploy to think of SQ.

It is recommended that employees and customers should share no difference in conceptual model and only partial difference in manifestation of constructs before managers can conclude that they have a similar SQ conceptualization (Cheung, 1999). The absence of these two forms of discrepancy symbolizes that employees and customers conceptualize SQ similarly. This is favorable to the organization as it reveals that employees and customers not only use the same SQ model to think of SQ, but also perceive the relative strength of relationships between the service components and the underlying SQ dimensions similarly.

Psychometric Properties of the Scale

Survey items are mostly executed to measure how respondents evaluate SQ. For example, in the case of SERVQUAL, there are altogether 22 items to measure SQ.

However, in the face of the scale items, employees and customers might respond to them differently. They might demonstrate differences in the psychometric properties of the measurement scale they use to evaluate SQ. There are three forms of discrepancy that manifest the two groups' differences in this area: *difference in magnitude of random error*, *difference in perception dispersion* and *difference in baseline perception.*

Difference in Magnitude of Random Error. In the ME/I literature, one of the tests is to examine the differences in the unique variances of items between two groups. As explained, unique variances are the error variances due to imperfection of measurement process. Thus, this test allows researchers to examine if employees and customers exhibit differences in the magnitude of random error when they are making SQ evaluation on the survey items.

An observed item score consists of a hypothetical true score plus random measurement error (Lord & Novick, 1968). In evaluating SQ (on the survey items), people might have random errors which are the variability of measures owing to random fluctuation with a characteristic of self-compensating (Kerlinger & Lee, 2000). Random errors are induced by random factors like mood of the respondents on the day they answer the survey questions.

Employees and customers might not show the same magnitude of random error in their SQ evaluation. Customers are usually regarded as more prone to exhibit more random errors because they might not be familiar with the service and therefore end up guessing on the SQ level of the service components mentioned in the survey items.

Also, customers might mix up the firm in question with other similar firms, experience lapses of memory, or be inattentive to the survey questions, all of which might engender in a larger magnitude of random error shown than employees would have.

Difference in Perception Dispersion. In examining the equivalence of factor variances between two groups in the ME/I literature, one's interest is in testing if there is any difference in the variability of the evaluated score on a factor between two groups (Vandenberg & Lance, 2000). Thus this test allows researchers and practitioners to examine if there is a difference in the degree of consensus (i.e., variability of the scores) about the SQ level of a SQ dimension between employees and customers.

A group's perception might disperse narrowly or widely. Perception dispersion denotes the degree of consensus about the SQ level within each group. There is a difference in perception dispersion when employees and employees do not have the same degree of consensus about the SQ level. In particular, when one group has *less* consensus regarding the SQ level of a particular dimension, they should show a larger perception dispersion comparing to another group.

If employees have larger SQ perception dispersion than customers do, that might be problematic. It gives the sign that employees have low concurrence regarding the SQ they have a part to provide on the SQ dimensions. One of the reasons might be role ambiguity, resulted from the lacking of service specification requirements given by management (Bruhn & Georgi, 2000). Without knowing the

standards their managers expect them to fulfill, each employee might create his/her own standards and therefore culminate in a wide range of perceived SQ levels among them. This might translate into larger perception dispersion when comparing to customers.

On the other hand, if customers have larger perception dispersion than employees do, again that might be a hitch. It reveals that customers do not quite concur on the perceived SQ level on a dimension the firm provides. One of the reasons might be the large fluctuation and low consistency in employees' performance, leading customers to have a wide range of perceived SQ levels. For instance, Hansen and Danaher (1999) found that customers experiencing inconsistent starts and finishes in a trend of service encounters gave more extreme evaluation of SQ, implying a possibility of larger perception dispersion than employees would have. Moreover, a strong culture or good training program might shape the way employees perceive the SQ levels. They might therefore show a larger consensus (smaller dispersion) about the SQ levels on the SQ dimensions when comparing to customers who are usually from a large variety of backgrounds.

Difference in Baseline Perception. In ME/I literature, testing the equality of intercept terms (i.e., scalar invariance) of items between two groups helps identify the differences in the starting points of SQ evaluations of the service components (mentioned in the items) between employees and customers.

The third form of discrepancy in psychometric properties of the measurement scale reflects in the fact that sometimes employees and customers might display

differences in their baseline perception of the SQ. Technically, baseline perception in itself denotes the perceived SQ level on the service component (mentioned in the survey item) when the latent dimension has zero effect. It is the "starting point" of the perceived SQ level on that service component. Thus, a difference in baseline perception between employees and customers reveals that one of the groups exhibits a higher starting point in the perceived SQ level, suggesting that they tend to be more lenient in giving SQ evaluation when comparing to another group (Cheung & Rensvold, 2000). For instance, customers might give a score of 1 on a 5-point Likert-type scale when they think the service is extremely poor whereas employees might still give a '2' to represent extremely poor performance as they might realize the difficulty of carrying out certain service components. Customers are more stringent in this case.

This form of discrepancy hints to managers on which service components customers have higher (lower) expectations, as manifested by stringency (leniency) in their SQ evaluations. Possible causes to induce differences in baseline perception might be the positive effect of good word-of-mouth, favorable consumer reports, or impressive advertisements (Zeithaml et al., 1993), all of which could boost customers' quality expectations. Therefore their SQ evaluations should become more stringent when comparing to employees, assuming that the positive impact of word-of-mouth, consumer reports, or advertisements is generally greater on customers than on employees.

SQ Evaluation

People incessantly evaluate or judge the SQ a firm provides. SQ evaluation usually manifests in the perceived SQ level, which is the judgment of the superiority or inferiority of the SQ (on a particular service component, or on a SQ dimension, or in general). It can be described as good (vs. bad) or high (vs. low). Thus, discrepancies in this area (SQ evaluation) emerge when employees and customers do not have the same perceived SQ level.

However, there is actually another form of SQ evaluation discrepancy that usually lacks of researchers' attention. It is the differences in the interrelationships among SQ dimensions. Employees and customers might perceive a different magnitude of relationship among several SQ dimensions in their SQ evaluations. These two forms of discrepancy are expounded below.

Difference in Interrelationships among Dimensions. This form of discrepancy corresponds to testing the equality of factor correlations in the ME/I literature where the interest is in examining whether two groups demonstrate differences in the perceived magnitude of interrelationships among the various dimensions.

Recall that there are usually several SQ dimensions in people's mindsets when they think of SQ (e.g., five dimensions in the case of SERVQUAL). Employees and customers might not perceive the same interrelationship among the SQ dimensions in their mindsets. Taking the SERVQUAL scale as an example. Empirical evidences show that these five SQ dimensions could be highly correlated (e.g., McDougall & Levesque, 1992). Therefore, it is reasonable to doubt that there might also be differences in the perceived magnitude of interrelationships between employees and

customers.

Possible reasons that might spur such differences in the perceived interrelationships among SQ dimensions between employees and customers might include employees' or customers' inability (Murphy & Jako, 1989) or unwillingness (Banks & Murphy, 1985) to discriminate among different SQ dimensions. For instance, customers might not be able to distinguish the dimension 'responsiveness' and 'assurance' (Parasuraman et al., 1991) and therefore perceive a higher interrelationship between these two dimensions in their perception vis-à-vis employees'.

Also, dominant dimensions influence the perception on other less important ones (Wirtz & Bateson, 1995). Since employees and customers might not agree on which particular dimension(s) is/are the dominant one(s), this might lead to different magnitude of perceived interrelationships among SQ dimensions. Other possible reasons might include customers' unfamiliarity with the service, or lack of the knowledge pertinent to the service, both of which lead them to rely on overall impressions of the service to evaluate each SQ dimension, resulting in a higher perceived interrelationship among SQ dimensions when comparing to employees (Leuthesser et al., 1995) who are usually more familiar and knowledgeable about the service.

The significance of examining this form of discrepancy between employees and customers lies in its implication for improvement of SQ. For example, if customers perceive a high relationship between the SQ dimensions "responsiveness" and

"assurance" and employees do not, there is a possibility that the "non-responsiveness" of employees might lead customers to judge the responsiveness aspect *plus* the assurance aspect of the SQ are poor while employees and managers might not realize it. Acknowledging the differences in the perceived interrelationships among SQ dimensions between the two groups is therefore important.

Difference in Perceived SQ level. In the ME/I literature, the testing of the equivalence of the level of a construct allows researchers to examine if there is any difference in the perceived SQ level of a construct between employees and customers. It is usually of huge interest for practitioners and researchers to know that whether employees and customers perceive the same SQ level on a SQ dimension because this form of discrepancy informs them whether employees over- or under-evaluate the SQ level when comparing to customers. This piece of information is enormously instrumental to management. If employees are found to have a significantly higher perceived SQ level on a dimension than customers do, it paraphrases that employees might think of themselves able to contently satisfy customers (on that SQ dimension) but in fact they did not. In other words, employees over-evaluated the SQ. The customers might eventually leave if no remedial actions are taken. Over-evaluation of the SQ by employees is not uncommon. For instance, when comparing to shippers (the customers), carriers (the employees) usually overestimate the SQ level they provided to the shippers (Hopkins et al., 1993). Similarly, hotel employees usually over-evaluate the SQ they think the hotel provides when comparing to hotel customers (McColl-Kennedy & White, 1997).

Possible reasons to bring differences in perceived SQ levels are again copious. For instance, those employees who are lack of feedback on their performance might over- or under-evaluate the SQ level; those employees who are lack of confidence might under-evaluate the SQ level they have a part to provide. Those who are of self-serving bias might not want to admit that the SQ is poor, and might therefore over-evaluate the SQ level. Also, when employees have a strong sense of pride working in a firm, they might over-evaluate the SQ level. It is absolutely to a firm's advantage if managers find that employees and customers share no differences in the perceived SQ levels on the SQ dimensions.

Summary

There can be seven forms of discrepancy existing between employees and customers which fall into three different areas. The first area concerns with those discrepancies related to *SQ conceptualization* which is the way people conceptualize or define SQ. The two forms of discrepancy in this area are the *difference in conceptual model* and *difference in manifestation of constructs*.

The second area concerns with those discrepancies pertinent to the *psychometric properties* of the measurement scale employees and customers use to evaluate the SQ. There are three particular forms of discrepancy under this area: *difference in magnitude of random error*, *difference in perception dispersion* and *difference in baseline perception*.

Finally, the third area concerns with those discrepancies germane to the *SQ*

evaluation. In particular, there are two forms of discrepancy in this area: *difference in interrelationships among dimensions* as well as *difference in perceived SQ level*.

The reasons suggested for each form of discrepancy certainly cannot be exhaustive. However, these reasons as well as the conceptual meaning of each form of discrepancy might guide managers to look for explanations if these forms of discrepancy are found. Equipped with such a more concrete understanding of the possible discrepancies existing between employees and customers, managers can accordingly educate their employees to fill the gaps.

CHAPTER V

METHODOLOGY

Data Collection

A dataset was collected for two purposes. First, it was used to examine whether any of the seven forms of discrepancy proposed in this study emerged between employees and customers. The second purpose was to illustrate how the seven forms of discrepancy could be identified using structural equations modeling (SEM), in particular, ME/I tests.

Survey Instrument. A survey was executed to both employees and customers in the gas station industry. Because of the long controversies about SERVQUAL's applicability and replicability, 31 items were generated mainly based on the SERVQUAL items for the current empirical investigation rather than directly adapting the 22-item SERVQUAL scale (Carman, 1990). These 31 items altogether served to measure respondents' evaluation of the SQ the gas stations provided. The items on the customer survey were as same as those on the employee survey. The survey was administered at 50 gas stations.

Regarding the operation of this 31-item scale, this study adopted the perspective of Cronin and Taylor (1992b) that SQ should be gauged using only performance measures, instead of the difference scores obtained from subtracting expectation scores from performance scores (Parasuraman et al., 1988). Usage of difference

scores in general results in low reliability when the two components are correlated (Peter, Churchill & Brown, 1993). Besides, there is increasing empirical evidence that SQ should be measured using performance-based measures only (e.g., Babakus & Boller, 1992; Bolton & Drew, 1991; Carman, 1990; Cronin & Taylor, 1992b; Durvasula et al., 1999; McDougall & Levesque, 1994).

Interview. Ten customers were interviewed at each of the 50 gas stations. In order to gain a more representative sample, the quota of 10 respondents was equally divided into 2 groups: private motorists and professional drivers. Only those who recounted the station as one of their three most frequently visited gas stations were qualified as appropriate respondents so as to gain fairer opinions. The final sample consisted of 242 professional drivers and 248 private motorists, adding up to 490 customers. The staff survey was executed during the day and the evening shifts. From each of the 50 gas stations, 8 staff were interviewed. The final sample consisted of 16 dealers, 83 station managers, 46 cashiers, and 211 station attendants, adding up to 356 employees. After elimination of cases with missing data using the listwise procedure, the effective sample size was 350 for employees and 447 for customers.

Method of Analysis

As mentioned earlier, in order to test the seven forms of discrepancy existing between employees and customers, SEM was used for analysis. Specifically, an array of measurement equivalence/invariance (ME/I) tests was executed.

Testing Difference in Conceptual Model. Employees and customers usually hold a SQ conceptual model in their mindsets. Testing the difference in conceptual model corresponds to testing the presence or absence of configural invariance in the ME/I literature. This test can be performed by fitting the same model separately to the data obtained from employees and customers. For example, a SERVQUAL model can be fitted to both groups to see if the model fits well. If it does, it means that a SERVQUAL model is a reasonable representation of how respondents conceptualize SQ.

It is known that a collection of indicators should be used to justify the model fitness instead of relying only on a single indicator (Vandenberg & Lance, 2000). Like Cheung (1999), overall fitness of the model was judged by using χ^2 likelihood ratio test, root mean square error of approximation (RMSEA; Steiger, 1990), non-normed index (TLI; Tucker & Lewis, 1973) and comparative fit index (CFI; Bentler, 1990). Since χ^2 test is sensitive to sample size, examining the mentioned fit indices is necessary for making valid inferences. Common practices usually take RMSEA less than .08 as acceptable whereas TLI and CFI should be at least .90.

If a theoretical model can have an acceptable fit to both employees and customers based on the aforementioned criteria, it can be concluded that there is no difference in conceptual model between employees and customers. This technically means that both employees and customers embody the same number of dimensions as well as the same pattern of item loadings to represent their SQ mindsets. They conceptualize SQ in a similar way. Examining the existence of other forms of discrepancy can be proceeded. If there is a difference in conceptual model, other tests

of discrepancy will not be meaningful because employees and customers are referring to two distinctive SQ conceptual models that are not directly comparable.

In order to ensure that the chosen conceptual model is a good representation of how the two groups define SQ, it is recommended to test other possible forms of conceptual models (Cheung, 1999). If these alternative models do not fit well to both groups at the same time, then researchers and practitioners could be more confident to conclude that the proposed/chosen model is concurrently used by employees and customers to define SQ.

Testing Difference in Manifestation of Constructs. A SQ dimension is manifested by items, and in particular, the relative strength of relationships between the items and the SQ dimensions. Testing the differences in manifestation of constructs between employees and customers corresponds to testing the presence of factorial invariance between two groups in the ME/I literature. But before going on to test whether there is such a difference, the two models above should be stacked together to form a baseline model in order to obtain a baseline χ^2 value for later comparisons.

Differences in manifestation of constructs are tested by constraining all factor loadings to be equal between two groups ($\lambda_{2,1}^{(c)} = \lambda_{2,1}^{(e)}$; $\lambda_{3,1}^{(c)} = \lambda_{3,1}^{(e)}$; ... with the superscript "c" and "e" represent customers and employees respectively). $\Delta\chi^2$ which is the difference between the χ^2 value of the constrained model and the baseline model (i.e., $\chi^2_{\text{constrained}} - \chi^2_{\text{baseline}}$) is used for testing. If $\Delta\chi^2$ is not significant, accompanied by an acceptable fit of the model, it will be inferred that the

constraint does not significantly worsen the fit of the baseline model. Therefore, no differences in manifestation of constructs are observed. On the other hand, if $\Delta\chi^2$ is significant plus worsening fit indices, it indicates that the model has a poorer fit than the previous unconstrained one. In other words, *full factorial invariance* does not exist. This inference is predicated on the fact that the model fits the data better if the factor loadings are not forced to be the same between employees and customers (Cheung & Rensvold, 2000).

When full factorial invariance does not exist, that is when two groups do not perceive the same strength of relationship between at least one item and the respective SQ dimension, item-level factorial invariance tests can be implemented to look for *partial factorial invariance* (Bryne et al., 1989). This operationally means that only one factor loading is constrained to be equal between employees and customers at one time, and see if the model fit will be significantly worsened. If not, that particular item is invariant between the two groups. After identifying the full/partial factorial invariance model, this model's χ^2 value can be used as the new comparison standard for subsequent tests.

Testing Difference in Magnitude of Random Error. In SEM, the error variance incurred in the measurement process can be estimated for each item. Thus, testing the differences in magnitude of random error between employees and customers corresponds to testing the equality of error variances in the ME/I literature.

In doing so, an equality constraint is placed on all pairs of error variances in the CFA model ($\theta_{1,1}^{(c)} = \theta_{1,1}^{(e)}$; $\theta_{2,2}^{(c)} = \theta_{2,2}^{(e)}$; ...). If the constraint brings a significant $\Delta\chi^2$

plus poorer fit indices, it indicates that the constraint is unwarranted. In other words, differences in magnitude of random error are found. Item-level tests can then be performed in order to identify which pair(s) of errors is/are of different values between the two groups. On the other hand, an insignificant $\Delta\chi^2$ and an acceptable fit means that there is no difference in magnitude of random error between the two groups.

Testing Difference in Perception Dispersion. Perception dispersion is the variability of the scores on a SQ dimension. It indicates the degree of consensus about the SQ level of that particular SQ dimension. In SEM, factor variance in essence represents the variability of the factor (Vandenberg & Lance, 2000). A large factor variance implies a smaller consensus about the SQ of a particular dimension within a group. Therefore, testing the equality of factor variances ($\phi_{1,1}^{(c)} = \phi_{1,1}^{(e)}$; $\phi_{2,2}^{(c)} = \phi_{2,2}^{(e)}$; ...) between two groups can help identify the differences in perception dispersion.

The equality constraint on factor variances is imposed on the CFA model. If the constraint brings a significant $\Delta\chi^2$ plus poorer fit indices, it indicates that differences in perception dispersion are found. Dimension-level tests can be carried out to check which dimension(s) employees and customers have different perception dispersion on. On the other hand, an insignificant $\Delta\chi^2$ and an acceptable fit means that there is no difference in perception dispersion.

Testing Difference in Baseline Perception. Baseline perception refers to the "starting point" of evaluation of the SQ level on a service component. In the ME/I

literature, testing the equality of items' intercepts terms (scalar invariance) between two groups helps test the existence of differences in baseline perception.

Differences in baseline perception can be identified by adding the equality constraint on the intercept terms between the two groups ($\tau_2^{(c)} = \tau_2^{(e)}$; $\tau_3^{(c)} = \tau_3^{(e)}$; ...). Again a significant $\Delta\chi^2$ and a poorer fit means that differences in baseline perception are found. Similar to other tests, item-level tests can be performed to identify which particular pair(s) of intercepts is/are different between the two groups.

Testing Difference in Interrelationships among Dimensions. As discussed, employees and customers might not perceive the same interrelationships among SQ dimensions. Differences in interrelationships among dimensions might be identified by examining the equality of factor correlations between two groups in SEM.

As mentioned before in the literature section, the result of testing the equality of factor covariances between two groups might be contaminated by the observed differences in factor variances. Thus, many researchers combine this test with the test of equality of factor variance into one simultaneous test in order to examine the equivalence of the factor correlations instead of factor covariances (Vandenberg & Lance, 2000). In doing so, Cheung (1999) suggests the following:

The correlation between construct 1 and construct 2 (ξ_1 and ξ_2) is $r_{12}^{(c)}$ for customers; and similarly $r_{12}^{(e)}$ for employees. The null hypothesis implies that $r_{12}^{(c)} = r_{12}^{(e)}$. Since

$$r_{12}^{(c)} = \frac{\phi_{12}^{(c)}}{\sqrt{\phi_{11}^{(c)}}\sqrt{\phi_{22}^{(c)}}} \quad \text{and} \quad r_{12}^{(e)} = \frac{\phi_{12}^{(e)}}{\sqrt{\phi_{11}^{(e)}}\sqrt{\phi_{22}^{(e)}}},$$

equating $r_{12}^{(c)} = r_{12}^{(e)}$ and solving for $\phi_{12}^{(e)}$ yields

$$\phi_{12}^{(e)} = \frac{\phi_{12}^{(c)}\sqrt{\phi_{11}^{(e)}}\sqrt{\phi_{22}^{(e)}}}{\sqrt{\phi_{11}^{(c)}}\sqrt{\phi_{22}^{(c)}}},$$

which is the appropriate form of constraint. This constraint can be added on each possible pair of factor correlations to the full/partial invariance model. A significant $\Delta\chi^2$ plus a poorer fit means that differences in interrelationships among dimensions exist. Again, only one pair of factor correlations is subsequently tested in order to identify those pairs of factors with different correlations between employees and customers.

Testing Difference in Perceived SQ level. This study purports to apply latent variable score (LVS), a new feature of LISREL (an application software to run SEM analysis), to compare the perceived SQ levels on the SQ dimensions between employees and customers instead of using latent means. As introduced in the ME/I tests review section, the level of the construct is usually compared using latent means. However, Cheung and Rensvold (2001) caution the following when making comparison between groups with a dimension's composite score:

There can be (at least) three kinds of composite score constructed for a dimension. They are summated average score (SAS), latent mean (LM) and latent variable score (LVS). SAS is generated by simply averaging the scores of the items representing a dimension. However, in doing so, it implicitly assumes that each item

has an equal importance to that dimension (i.e., unit weight), an assumption that is often incorrect (Reise, Widaman & Pugh, 1993). It is common that some items are more imperative to respondents and some items are less, even under the same dimension. In that case, it is not justifiable to simply average the item scores. The second problem of using SAS is that it ignores the differences two parties perceive in the relative strength of relationship between the items and the underlying dimensions. For example, customers might perceive a close relationship between the item "XYZ's physical facilities are visually appealing" and the dimension "tangibles" whereas employees might think that this item is not much related to "tangibles" at all. If SAS is used to compare the perceived SQ levels on the dimension "tangibles" between the two groups, this difference will be ignored and the result might not be fully correct.

In solving the first problem of using SAS, SEM allows the estimation of a LM which is an estimated mean score for a factor (dimension) that takes into account of different factor loadings (relative strength of relationships) of the items. Also, like other parameters estimated in SEM, the variances due to random error are partialled out in the estimation of LMs. Thus, using LMs is better than using SAS because using LMs does not assume the same strength of relationships between items and the underlying dimensions and allows for the de-attenuation of error variances.

However the second problem of using SAS remains that comparison between two groups using LMs necessitates a technical requirement imposing on the indicators: they have to show *both* factorial invariance as well as intercept (scalar) invariance between the two groups before they can be retained in the LMs

comparison. If any item is non-invariant *either* in the factor loading or intercept between the two groups, it will be effectively dropped out of the LMs comparison, and the result again might not be able to reflect the truest picture. Hence, Cheung and Rensvold (2001) propose to compare the level of the constructs by LVS which is a newly invented feature of LISREL 8.30 (Jöreskog, 2000). It is in essence a factor score and is estimated for each particular case. There are several properties of LVS (Jöreskog, 2000):

- The LVS, like any other parameter estimated using SEM, is deattenuated for variances due to measurement errors (unique variances of items).
- The scores are unbiased estimates of the latent variables
- The covariance matrix of the LVS equals the estimated variance-covariance matrix of the latent variables.

More importantly, LVS takes into account the non-invariant factor loadings *and* non-invariant intercepts into its estimation. In other words, while LMs comparison requires the presence of *both* factorial invariance and intercept (scalar) invariance, LVS comparison requires neither. Given these merits of LVS, this study applies LVS to test the differences in the perceived SQ levels instead of LM (especially when it is always difficult to get *all* factor loadings *and* intercepts to be equal between two groups). Dissemination of this piece of information regarding the merits of LVS should be conducive to all those research that involve between-group comparisons using survey-items (Cheung & Rensvold, 2001).

This study therefore takes Cheung & Rensvold's (2001) recommendation to test

the difference in the level of a construct between employees and customers using LVS. Nevertheless there is still no invariance test for LVS in the application software, LISREL. Therefore, after obtaining the LVS for each respondent, an independent sample t-test will be performed for each pair of dimensions. A significant difference suggests that employees and customers do not agree on the quality of that particular dimension the firm provides. Employees might have over- or under-evaluated the SQ level of that dimension.

Other Methodologies in Examining ME/I

Regarding other feasible methodologies in examining ME/I, Cheung (1999) has provided an excellent discussion in his study which aims at documenting several forms of self-others rating disagreement using ME/I tests. As the author implies, ME/I tests have advantages over these methods and therefore should be used instead. Some of these alternative methods are introduced below, accompanied by their possible drawbacks.

A difference in factor forms (i.e., configural invariance) can be tested by visually comparing the factor structures obtained from an exploratory factor analysis (EFA) for the two groups. Comparisons can be made on the number of dimensions obtained for each group, and also on the pattern of the item loadings. However, EFA does not render a test for statistical significance between the two factor structures and therefore the result might be at the mercy of subjectivity. Besides, there are various ways to extract the factors in EFA, culminating in different possible factor structures even within the same group (Bollen, 1989).

Differences in factor loadings (i.e., factorial invariance) across groups can be tested by comparing the correlations among factor loadings across groups. However, Cheung (1999) warns that comparing the correlations among factor loadings requires a large number of items and the method cannot be applied to comparisons more than two groups.

Differences in measurement errors can be tested by comparing the Cronbach's alpha across groups. However, currently there is no statistical test for the significance of a difference in Cronbach's alpha (Mullen, 1995). Thus, the interpretation of the magnitude of difference is again subjective.

Differences in the factor variances can be examined by comparing the standard deviations of each dimension's items set across groups. However, again there is no statistical significance test for the difference in standard deviations up to date.

Differences in the intercept terms of items (i.e., scalar invariance) can be tested by a profile analysis (Morris & Pavett, 1992). Specifically, when plotting the means of constructs for the two groups in the same graph, the lack of parallelism of the two lines "suggests that the differences between the two data sets are not caused by systematic response bias". However, even if the lines in the graph are really parallel, the differences in means may not all due to response bias (Mullen, 1995). Thus, this method might give rise confusion.

Differences in the interrelationships among dimensions can be compared by

inspection. With Fisher's z transformation, researchers can test the significance of a difference between the two correlations. However, the correlations are contaminated with measurement error.

Finally, as mentioned, a difference in the level of a dimension between two groups can be tested by comparing a dimension's composite score. There are at least three kinds of composite score for a dimension: summated average score (SAS), latent mean (LM) or latent variable score (LVS). LVS should be a better choice though (see the section of Method of Analysis).

However, researchers in the arena of examining the differences in employees' and customers' perceived SQ levels seldom used composite scores to make comparison. Rather, they usually contrasted them by executing multiple item-level comparisons (e.g., Brown & Swartz, 1989; Comila, 2000; Hopkins et al., 1993; McColl-Kennedy & White, 1997). In other words, they get insight on the differences in the perceived SQ levels between the two parties by comparing the evaluation of SQ level on each survey item.

The procedure typically entails: first, respondents, both employees and customers, evaluate the SQ level of the service components provided by a firm on a range of survey items. It can be construed that these items as a whole represent the more abstract concept "service quality" (e.g., all 20 items are constructed to represent SQ). Then researchers compare the mean score of each pair of same statements evaluated by the two groups, using t-test. A significant result demonstrates that employees and customers do not have the same perceived SQ level on the service

component mentioned in that particular statement. If significant differences are found on most statements, it suggests that employees and customers do not have a similar perceived SQ level in general.

Comparing the mean score of each pair of same statements has an advantage of allowing researchers to know which particular facet(s) of service (i.e., service components) is/are evaluated differently. For instance, if employees are found to have a significantly higher score on the statement "the employees in the firm XYZ are trustable", this conveys information to managers that whereas employees regard they are trustable, customers do not. Accordingly, managers can focus in improving the trustworthiness of their employees. In this fashion, comparing each pair of evaluated statements delivers information about the specific differences existing between two groups.

However, for a 20- or even 30-statement survey, making 30 t-tests can be very time-consuming and tedious. Second, when executing multiple significance tests, researchers are prone to commit type I error if they do not have corresponding adjustment on the chosen significance level. Third, the statement scores are tainted with measurement error. Fourth, individual item scores do not convey as much information as composite scores do (Kim & Mueller, 1994; Mullen, 1995). Finally, there is a possibility that out of all evaluated statements, half or nearly half of them are found to be differently evaluated while another half are not. If that is the case, a conclusion is difficult to make on whether the two parties are having a similar or dissimilar perceived SQ level in general.

Being convenient to use, correlation is also usually deployed to make comparisons between employees' and customers' perceived SQ levels in the SQ literature. When two parties' perceived SQ levels are positively correlated (e.g., Klose & Finkle, 1995; Reynierse & Harker, 1992; Schneider & Bowen, 1985; Schneider et al., 1980), it can be inferred that they have a similar SQ judgment because when one party sees the SQ level as good, it is likely that another party also does. By the same token, when two parties' perceived SQ levels are negatively correlated, it can be inferred that they have a dissimilar perceived SQ level. However, although correlation coefficient is convenient to use, it does not directly pinpoint the difference or equality of the perceived SQ levels between the two parties. Their perceived SQ levels might be correlated, but are they close? Unfortunately, using correlation cannot tell. Besides, correlation coefficients are again affected by measurement errors.

CHAPTER VI

RESULTS

The means, standard deviations and the correlation matrix of the 31 items for employees and customers were given in Table 4 and Table 5 respectively. The analysis result of each form of discrepancy in the area of SQ conceptualization, psychometric properties of the measurement scale and SQ evaluation would be explicated next.

Table 4

Means, Standard Deviations, and Correlations of Employees Data

Item	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	5.62	1.40	1																														
2	6.29	.94	.34	1																													
3	6.43	.87	.21	.28	1																												
4	6.44	.79	.16	.26	.36	1																											
5	5.98	1.04	.15	.26	.36	.33	1																										
6	6.51	.76	.26	.28	.43	.49	.35	1																									
7	6.26	.91	.22	.27	.32	.46	.35	.42	1																								
8	6.13	.98	.31	.31	.33	.40	.35	.50	.56	1																							
9	6.23	1.00	.22	.27	.38	.51	.35	.51	.50	.48	1																						
10	6.32	.86	.31	.28	.37	.48	.36	.47	.54	.53	.56	1																					
11	5.73	1.43	.24	.22	.16	.10	.14	.22	.24	.27	.17	.21	1																				
12	6.44	.77	.27	.26	.37	.38	.27	.43	.44	.45	.41	.52	.30	1																			
13	5.94	1.15	.24	.23	.23	.31	.32	.39	.33	.34	.45	.38	.20	.38	1																		
14	6.67	.69	.19	.31	.16	.19	.13	.21	.26	.24	.22	.32	.27	.33	.35	1																	
15	5.58	1.17	.28	.30	.28	.25	.32	.33	.36	.39	.22	.41	.20	.35	.29	.22	1																
16	5.87	1.03	.24	.27	.31	.34	.34	.35	.38	.43	.38	.39	.28	.39	.32	.32	.47	1															
17	6.53	.77	.25	.26	.53	.44	.39	.55	.44	.50	.55	.58	.19	.44	.40	.36	.36	.36	1														
18	6.25	.88	.23	.22	.34	.38	.38	.46	.46	.47	.50	.50	.25	.50	.39	.29	.42	.45	.51	1													
19	5.91	1.07	.18	.12	.24	.32	.28	.36	.34	.32	.38	.32	.13	.33	.43	.22	.24	.38	.30	.37	1												
20	6.42	.77	.22	.26	.39	.51	.35	.44	.42	.44	.48	.58	.18	.53	.34	.33	.32	.39	.51	.48	.37	1											
21	6.42	.76	.22	.23	.35	.47	.33	.40	.49	.46	.40	.51	.13	.37	.39	.43	.36	.41	.47	.45	.36	.49	1										
22	6.45	.79	.29	.30	.44	.44	.39	.56	.45	.55	.58	.63	.20	.55	.42	.33	.37	.41	.71	.57	.40	.63	.54	1									
23	5.79	1.06	.30	.16	.31	.37	.34	.42	.47	.51	.47	.51	.25	.43	.32	.22	.45	.46	.42	.45	.34	.43	.43	.56	1								
24	6.14	.99	.17	.18	.29	.41	.31	.40	.46	.50	.53	.49	.17	.43	.45	.33	.38	.51	.48	.49	.47	.48	.45	.61	.56	1							
25	6.19	.89	.20	.25	.29	.43	.42	.47	.55	.53	.49	.51	.28	.45	.42	.28	.32	.43	.45	.51	.40	.51	.49	.59	.49	.54	1						
26	6.57	.72	.21	.26	.44	.49	.33	.46	.46	.40	.44	.47	.13	.44	.50	.38	.23	.22	.49	.44	.39	.44	.54	.60	.41	.48	.49	1					
27	6.12	.92	.28	.26	.31	.42	.42	.44	.45	.52	.45	.53	.28	.41	.40	.31	.32	.48	.43	.48	.39	.55	.43	.56	.49	.49	.61	.47	1				
28	5.88	1.07	.16	.19	.24	.30	.32	.31	.38	.39	.39	.43	.19	.45	.44	.26	.29	.37	.37	.31	.46	.43	.35	.44	.43	.43	.45	.41	.46	1			
29	6.08	1.14	.17	.20	.28	.27	.25	.35	.34	.33	.35	.27	.43	.30	.37	.30	.20	.28	.37	.28	.30	.31	.33	.38	.32	.30	.39	.43	.35	.44	1		
30	6.08	1.10	.17	.19	.32	.41	.24	.35	.32	.26	.35	.32	.19	.37	.30	.30	.26	.31	.37	.33	.30	.39	.45	.41	.44	.37	.41	.51	.38	.39	.41	1	
31	6.41	.82	.22	.27	.37	.56	.32	.50	.53	.55	.49	.56	.23	.51	.34	.29	.31	.40	.56	.54	.39	.49	.53	.60	.52	.54	.54	.56	.49	.38	.42	.49	1

The item numbers correspond to the items appearing on the survey. See Appendix 1 for the survey items.

Table 5

Means, Standard Deviations, and Correlations of Customers Data

Item	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	5.20	1.24	1																														
2	4.95	1.33	.46	1																													
3	5.47	1.10	.42	.39	1																												
4	5.41	1.15	.52	.49	.60	1																											
5	5.06	1.24	.43	.44	.54	.53	1																										
6	5.37	1.18	.55	.51	.56	.64	.63	1																									
7	5.70	1.06	.47	.49	.62	.57	.51	.66	1																								
8	5.14	1.22	.45	.47	.53	.53	.55	.58	.57	1																							
9	5.40	1.14	.49	.43	.55	.46	.53	.60	.60	.60	1																						
10	5.52	1.13	.48	.49	.55	.56	.56	.62	.65	.58	.67	1																					
11	5.52	1.36	.32	.31	.36	.41	.40	.41	.41	.36	.42	.44	1																				
12	5.43	1.17	.45	.45	.55	.55	.52	.62	.61	.60	.62	.64	.44	1																			
13	5.01	1.27	.33	.28	.46	.39	.42	.43	.41	.34	.42	.42	.46	.47	1																		
14	6.04	.99	.26	.27	.38	.34	.28	.40	.47	.30	.44	.42	.48	.39	.42	1																	
15	4.91	1.20	.42	.45	.47	.49	.51	.52	.46	.52	.57	.54	.37	.60	.40	.34	1																
16	5.11	1.10	.37	.42	.43	.40	.46	.51	.47	.44	.53	.55	.34	.56	.45	.38	.64	1															
17	5.58	1.11	.50	.42	.63	.59	.54	.64	.63	.56	.64	.66	.45	.66	.45	.44	.57	.52	1														
18	5.06	1.19	.42	.40	.49	.51	.54	.57	.47	.57	.49	.53	.36	.59	.42	.26	.61	.51	.56	1													
19	5.13	1.21	.22	.25	.42	.38	.29	.35	.37	.30	.39	.34	.36	.42	.51	.34	.43	.38	.41	.38	1												
20	5.62	1.04	.49	.44	.53	.61	.43	.59	.61	.50	.55	.55	.46	.56	.42	.43	.49	.49	.65	.52	.46	1											
21	5.50	1.16	.44	.42	.46	.61	.42	.55	.53	.41	.51	.52	.49	.52	.41	34	.48	.44	.59	.50	.44	.66	1										
22	5.56	1.10	.47	.45	.62	.62	.52	.61	.62	.55	.67	.66	.44	.63	.46	.41	.56	.50	.75	.60	.43	.63	.60	1									
23	4.89	1.14	.42	.40	.42	.48	.50	.53	.44	.53	.49	.55	.33	.52	.42	.20	.59	.48	.52	.64	.36	.51	.46	.59	1								
24	5.20	1.15	.39	.46	.53	.51	.37	.55	.55	.50	.51	.55	.36	.54	.45	.35	.53	.42	.57	.50	.39	.52	.50	.58	.56	1							
25	5.38	1.14	.43	.44	.60	.58	.60	.62	.58	.53	.60	.64	.42	.62	.45	.36	.57	.56	.65	.60	.41	.58	.58	.67	.62	.63	1						
26	5.57	1.10	.29	.31	.52	.44	.44	.42	.45	.38	.51	.52	.38	.53	.48	.40	.42	.44	.58	.43	.45	.51	.52	.57	.42	.49	.59	1					
27	5.46	1.09	.40	.41	.53	.53	.44	.53	.57	.46	.56	.57	.40	.59	.45	.37	.52	.56	.60	.50	.43	.60	.58	.61	.53	.61	.65	.60	1				
28	5.34	1.13	.26	.36	.42	.36	.39	.39	.42	.36	.41	.41	.39	.47	.56	.35	.47	.49	.41	.41	.47	.44	.46	.43	.41	.53	.49	.57	.55	1			
29	5.47	1.33	.30	.29	.34	.43	.35	.35	.38	.27	.34	.38	.51	.38	.45	.34	.36	.35	.35	.35	.34	.44	.52	.39	.39	.36	.45	.42	.43	.52	1		
30	5.38	1.22	.46	.42	.48	.54	.47	.50	.49	.43	.52	.51	.46	.56	.40	.31	.50	.46	.56	.51	.45	.58	.62	.57	.56	.56	.60	.48	.59	.49	.49	1	
31	5.52	1.10	.40	.41	.55	.50	.46	.53	.56	.46	.53	.54	.43	.58	.47	.38	.52	.48	.59	.52	.47	.57	.57	.61	.56	.61	.66	.55	.69	.50	.47	.65	1

The item numbers correspond to the items appearing on the survey. See Appendix 1 for the survey items.

SQ Conceptualization

Difference in Conceptual Model. To begin, the existence of difference in conceptual model was tested by separately fitting the same model to both groups. As narrated, the fit was evaluated by χ^2, RMSEA, TLI and CFI. The significance level for rejection was chosen to be at $p<0.01$ as multiple significant tests were carried out. The first indicator of each construct was chosen to be the marker item. Their factor loadings were set to be equal to one for identification purposes.

As mentioned earlier, several possible alternative models should be tested in order to identify the model that could best represent employees' and customers' SQ conceptualization. In the current study, three models, which were suggested in the SQ literature, would be tested. The first one was a SERVQUAL model. The 31 items executed in the survey were generated mainly based on SERVQUAL items. If SERVQUAL was a good representation of how people conceptualize SQ, these 31 items should also show a similar 5-factor structure as that of the SERVQUAL. These 31 items thus were carefully matched with the five SERVQUAL dimensions based on the content of the items as well as the meaning of each dimension. The five dimensions now were represented by five to seven items. The pattern of items representing these five dimensions was given in Appendix 2. This 31-item 5-factor SERVQUAL model would be tested for the fitness.

From the literature, it was found that the correlations among the SERVQUAL dimensions could be quite high, implying that there might be fewer dimensions than the SERVQUAL suggested. For instance, Cronin and Taylor (1992a) failed to

replicate the 5-factor structure after administering the 22-item SERVQUAL scale to respondents. They instead found a unidimensional structure using these 22 items in four service industries. People might sometimes just form an overall unidimensional abstraction of SQ (Babakus & Boller, 1992). Therefore a 31-item unidimensional factor structure would also be tested in the current study as the second alternative model.

The third alternative model was suggested by Mels et al. (1997). In particular, they proposed that there were two dimensions of SQ. The first dimension, which was called "extrinsic SQ", was composed of those SERVQUAL items about tangibles. Extrinsic SQ refers to "what is used during the service delivery process". It is similar to the SQ dimension "physical quality" proposed by Lehiten and Lehiten (1985). The second dimension was called "intrinsic SQ" and was composed of the rest of the SERVQUAL items. The authors suggested that intrinsic SQ mainly refers to the attributes and actions of employees (e.g., courtesy, knowledge and promptness). This dimension corresponds to the SQ dimension "interactive quality" proposed by Lehtinen and Lehtinen (1985).

The tenability of such a 2-factor model was also supported in a review of literature. First, there were evidences that the dimension "tangibles" often remained in most SQ conceptual models (Llosa et al., 1998). Besides, except for the dimension tangibles, the rest of the four SERVQUAL dimensions were found to be highly correlated from .52 to .81. (McDougall & Levesque, 1994). It implied that these four dimensions might be combined to form one dimension. Spreng and Singh (1993) also hinted at the possible combination of some of the five dimensions of SERVQUAL.

Thus, this 2-factor model proposed by Mels et al. (1997) would also be tested as the third alternative SQ model.

An examination of the 31 items in an attempt to divide them into "intrinsic SQ" items and "extrinsic SQ" items however showed that these items could not all be categorized into these two areas. Specifically, seven items concerned with the physical facilities of the gas stations (extrinsic SQ). For example, one of the items was "the gas stations have up-to-date equipment". Seventeen items concerned with the actions and attributes of employees (intrinsic SQ). One of the items as an example was "the employees are polite". The remaining seven items did not clearly describe the quality of either the physical facilities or the attributes and actions of employees. An example of those items was "the gas station knows what the needs of the customers are". It was ambiguous in that respondents might not know whether it was referring to the quality of the facilities in satisfying the customers' needs or the employees' concern in customers' needs. Consequently, these seven items were eliminated. Therefore, for this alternative model, there were altogether 24 items in total. The pattern of items representing these two dimensions was given in Appendix 2.

The result of testing the fitness of these three models was given in table 6. The 31-item SERVQUAL model as well as the 31-item unidimensional model could not be adequately fitted to both employees and customers. For employees data, the poor fit of these two models was manifested in the values of TLI and CFI which range from .85 to .87. For customers data, the poor fit of these two models was manifested in the values of RMSEA which were larger than .08, as well as in the values of TLI

and CFI which were smaller than .90.

TABLE 6

Results on Testing Differences in Conceptual Model

Model	χ^2	df	RMSEA	TLI	CFI
		Employees			
SERVQUAL	1119.51	424	.069	.853	.866
Unidimensional	1170.56	434	.070	.848	.858
2-factor (24 items)	598.98	251	.062	.896	.905
2-factor (23 items)	499.96	229	.058	.912	.920
		Customers			
SERVQUAL	1514.62	424	.086	.871	.882
Unidimensional	1611.63	434	.089	.864	.873
2-factor (24 items)	763.99	251	.072	.917	.924
2-factor (23 items)	670.44	229	.071	.924	.932

On the other hand, the 24-item 2-factor model fitted both groups well (Table 6). However, an examination of modification indices in this model indicated that the model fit could be significantly improved by allowing one item (the gas stations can always keep the area clean and neat) for the dimension "extrinsic SQ" to cross-load on the dimension "intrinsic SQ" (the modification index showed that the chi-square value would be reduced by 10.98 for employees and 13.71 for customers if the item was allowed to load on the intrinsic SQ dimension). It suggested that this particular item did not clearly load on either factor in both groups. Thus, this item was deleted and the final model consisted of 6 items measuring "extrinsic SQ" and 17 items measuring "intrinsic SQ" (see Appendix 2 for the items' pattern of this model).

The fitness for this 23-item 2-factor model was again shown in table 6. This model fitted to both employees and customers data well. Specifically, the χ^2 value was 499.96 (df=229), RMSEA was .058, TLI was .912 and CFI was .920 for employees data. For customers data, the χ^2 value was 670.44 (df=229), RMSEA was .071, TLI was .924 and CFI was .932. It showed that this 2-factor model could be an adequate representation of how employees and customers conceptualized SQ in their mindsets. Both groups embodied two dimensions (intrinsic and extrinsic SQ) and associated the same patterns of items to think of these two SQ dimensions. This convergence gave us confidence to carry on testing the existence of other forms of discrepancy.

Before proceeding to other tests, it was better to examine the discriminant validity of these two constructs in the respondents' mindsets as the correlations between SERVQUAL dimensions were found to be quite high (e.g., Babakus & Boller, 1992). Since the 31 survey items in the current study were generated mainly based on SERVQUAL items, it was reasonable to doubt that these two dimensions, extrinsic and intrinsic SQ, were highly correlated.

The factor correlation between these two constructs for employees and customers was .808 and .782 respectively. The factor correlation between these two constructs was tested against a value of one for each group. The result was shown in Table 7. It could be seen that when the value of the factor correlation was constrained to be a value of one in the original model, the $\Delta\chi^2$ change was significant plus model fit became worse when comparing to the original unconstrained model for both employees and customers. In other words, the constraint was unwarranted. This

suggested that the factor correlation between these two dimensions was significantly different from a value of one in both groups. It was safe to conclude that these two constructs (extrinsic and intrinsic SQ) were perceived as two distinctive constructs in both groups. This final model would be adopted to carry on to other tests of discrepancy. It was depicted in Figure 1. It should be noted that the choice of the maker items in this model was not at all arbitrary. Using Cheung and Rensvold's (1999) iterative procedure, various sets of non-invariant items were identified. Inasmuch as managers are usually interested in identifying the differences between employees and customers, the set of marker items that associates with the most non-invariant items was chosen for this model.

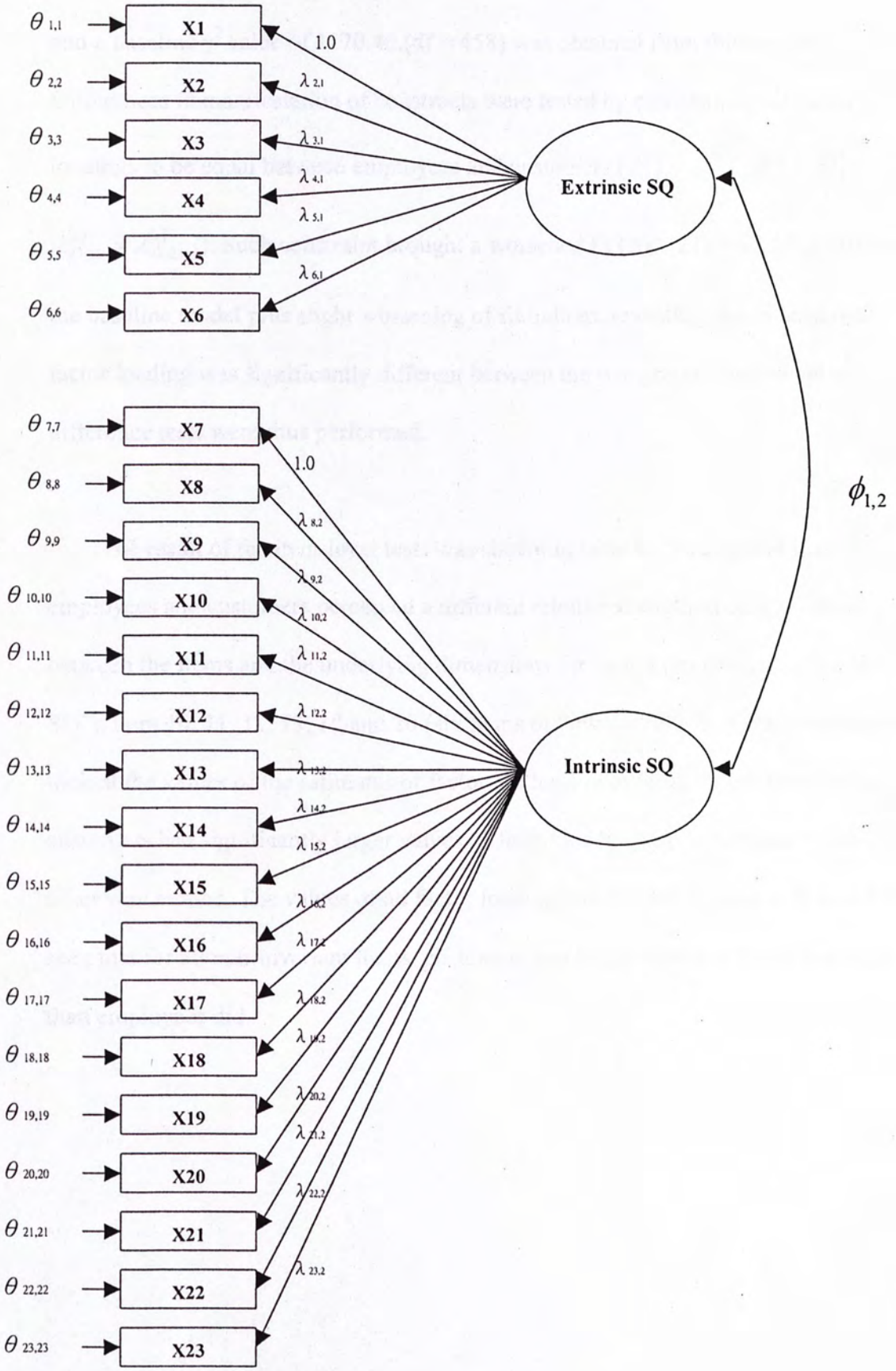
TABLE 7

Results on Testing the Factor Correlations Against a Value of One

	X^2	df	$\Delta\chi^2$	Δdf	p-value	RMSEA	TLI	CFI
Employees	576.28	230	76.32	1	.000	.067	.889	.898
Customers	886.71	230	216.27	1	.000	.089	.888	.898

Figure 1

The Final Model



Difference in Manifestation of Constructs. The next step was to establish a baseline model for subsequent comparisons. The two models were stacked together and a baseline χ^2 value of 1170.40 (df = 458) was obtained from this model. Differences in manifestation of constructs were tested by constraining all factor loadings to be equal between employees and customers ($\lambda_{2,1}^{(c)} = \lambda_{2,1}^{(e)}$; $\lambda_{3,1}^{(c)} = \lambda_{3,1}^{(e)}$; ... ; $\lambda_{23,23}^{(c)} = \lambda_{23,23}^{(e)}$). Such constraint brought a worsened fit ($\Delta\chi^2$ (21) = 63.45, p<.01) to the baseline model plus slight worsening of fit indices, revealing that at least one factor loading was significantly different between the two groups. Item-level χ^2 difference tests were thus performed.

The result of the item-level tests was shown in table 8. It was found that employees and customers perceived a different relative strength of relationships between the items and the underlying dimensions for item 5 (an item of "extrinsic SQ"), item 10, 11, 12, 13, 18 and 20 (six items of "intrinsic SQ"). It was necessary to look at the values of the estimates of factor loadings in order to determine whether customers had significantly larger values of factor loadings in these items or the other way around. The values of all factor loadings were given in table 9. It could be seen that for all non-invariant items, customers had larger values of factor loadings than employees did.

TABLE 8

Results on Testing Differences in Manifestation of Constructs

Constraints	X^2	df	$\Delta\chi^2$	Δdf	p-value	RMSEA	TLI	CFI
Stacked	1170.40	458				.066	.920	.928
ALL	1233.85	479	63.45	21	.000*	.066	.919	.923
LX 2 1[a]	1171.46	459	1.06	1	.303	.066	.920	.928
LX 3 1	1175.96	459	5.56	1	.018	.066	.920	.927
LX 4 1	1171.38	459	.98	1	.322	.066	.920	.928
LX 5 1	1179.85	459	9.45	1	.002*	.066	.919	.927
LX 6 1	1170.72	459	.32	1	.572	.066	.920	.928
LX 8 2	1176.02	459	5.62	1	.018	.066	.920	.927
LX 9 2	1173.45	459	3.05	1	.081	.066	.920	.927
LX 10 2	1182.18	459	11.78	1	.001*	.066	.919	.927
LX 11 2	1185.70	459	15.30	1	.000*	.066	.919	.926
LX 12 2	1178.91	459	8.51	1	.004*	.066	.919	.927
LX 13 2	1189.08	459	18.68	1	.000*	.066	.918	.926
LX 14 2	1172.95	459	2.55	1	.110	.066	.920	.927
LX 15 2	1172.01	459	1.61	1	.204	.066	.920	.928
LX 16 2	1171.39	459	.99	1	.320	.066	.920	.928
LX 17 2	1174.54	459	4.14	1	.042	.066	.920	.927
LX 18 2	1189.05	459	18.65	1	.000*	.066	.918	.926
LX 19 2	1174.28	459	3.88	1	.049	.066	.920	.927
LX 20 2	1182.38	459	11.98	1	.001*	.066	.920	.927
LX 21 2	1175.03	459	4.63	1	.031	.066	.920	.927
LX 22 2	1174.82	459	4.42	1	.036	.066	.920	.927
LX 23 2	1175.90	459	5.50	1	.019	.066	.920	.927
PFI	1186.23	472	15.83	14	.324	.065	.922	.927

*$p < .01$

[a] LX is the LISREL syntax for factor loading of exogenous variable

TABLE 9

Values of Factor Loadings

Factor Loadings	Employees	Customers
LX (1,1) [a]	1.0	1.0
LX (2,1)	.9441	1.0796
LX (3,1)	.7498	1.1167
LX (4,1)	1.0335	1.1635
LX (5,1)	**.4446**	**.7195**
LX (6,1)	.8725	.9401
LX (7,2)	1.0	1.0
LX (8,2)	.9000	1.1541
LX (9,2)	.7187	.9610
LX (10,2)	**.6611**	**1.0163**
LX (11,2)	**.6871**	**1.0865**
LX (12,2)	**.7569**	**1.1043**
LX (13,2)	**.7361**	**1.1830**
LX (14,2)	.8598	1.0251
LX (15,2)	.9684	1.1126
LX (16,2)	1.0069	1.1186
LX (17,2)	.9277	1.1410
LX (18,2)	**.7228**	**1.1688**
LX (19,2)	.8305	1.0743
LX (20,2)	**.8112**	**1.1603**
LX (21,2)	.8543	1.0840
LX (22,2)	.7802	.9797
LX (23,2)	.9189	1.1540

[a] LX is the LISREL syntax for factor loading of exogenous variable

(Those bolded figures are non-invariant items)

Partial factorial invariance (PFI) was then invoked by freeing out these seven parameters in the stacked model. From table 8, it could be seen that the PFI model had a χ^2 value of 1186.23 (df = 472). This model became the new comparison standard for later analyses.

Psychometric Properties of the Scale

Difference in Magnitude of Random Error. Next, differences in magnitude of random error were tested by constraining all corresponding pairs of error variances to

be equal between the two groups ($\theta_{1,1}^{(c)} = \theta_{1,1}^{(e)}$; $\theta_{2,2}^{(c)} = \theta_{2,2}^{(e)}$; ... ; $\theta_{23,23}^{(c)} = \theta_{23,23}^{(e)}$). Such constraint brought a significant $\Delta\chi^2$ change ($\Delta\chi^2$ (23) = 286.63, p<.01) plus slight worsening of fit indices. Consequently, the equality constraint was imposed on only one pair of error variances at one time so as to identify those pairs with unequal error variances. The result was shown in table 10. It indicated that there was a significant difference between the two groups in the error variances on item 3, 5 (two items of "extrinsic SQ"), 9, 11, 13, 15, 17 to 23 (eleven items of "intrinsic SQ").

Again, the estimates were checked in order to know which group possessed larger values of unique variances in these identified items. The values of all error variances were given in table 11. It was found that customers possessed a significantly larger random error variance in their SQ evaluation on item 5, 11, 13, 15, 17, 18, 20, 21, 22 and 23 whereas employees showed a larger random error variance on item 3, 9 and 19 only.

TABLE 10

Results on Testing Differences in Magnitude of Random Error

Constraints	χ^2	df	$\Delta\chi^2$	Δdf	p-value	RMSEA	TLI	CFI
PFI	1186.23	472				.065	.922	.927
ALL	1454.86	495	268.63	23	.000*	.073	.900	.902
TD 1 1[a]	1186.25	473	.02	1	.888	.065	.922	.928
TD 2 2	1190.08	473	3.85	1	.050	.065	.922	.927
TD 3 3	1205.43	473	19.20	1	.000*	.066	.920	.926
TD 4 4	1186.24	473	.01	1	.920	.069	.922	.928
TD 5 5	1212.81	473	26.58	1	.000*	.066	.920	.925
TD 6 6	1188.79	473	2.56	1	.110	.069	.922	.927
TD 7 7	1186.42	473	.19	1	.663	.065	.922	.928
TD 8 8	1188.66	473	2.43	1	.119	.065	.922	.927
TD 9 9	1215.48	473	29.25	1	.000*	.066	.919	.925
TD 10 10	1186.39	473	.16	1	.689	.065	.922	.928
TD 11 11	1200.32	473	14.09	1	.000*	.065	.921	.926
TD 12 12	1186.23	473	.00	1	1	.069	.922	.928
TD 13 13	1207.56	473	21.33	1	.000*	.070	.920	.925
TD 14 14	1186.24	473	.01	1	.920	.065	.922	.928
TD 15 15	1197.78	473	11.55	1	.001*	.065	.921	.926
TD 16 16	1186.26	473	.03	1	.862	.065	.922	.928
TD 17 17	1201.11	473	14.88	1	.000*	.065	.921	.926
TD 18 18	1203.28	473	17.05	1	.000*	.065	.921	.926
TD 19 19	1196.66	473	10.43	1	.001*	.069	.921	.926
TD 20 20	1200.04	473	13.81	1	.000*	.065	.921	.926
TD 21 21	1210.25	473	24.02	1	.000*	.066	.920	.925
TD 22 22	1206.85	473	20.62	1	.000*	.065	.920	.925
TD 23 23	1216.72	473	30.49	1	.000*	.065	.919	.924

*$p < .01$

[a] TD is the LISREL syntax for theta-delta

TABLE 11

Values of Variances of Random Error

Variance of Random errors	Employees	Customers
TD 1 1[a]	.6372	.6263
TD 2 2	.8226	1.0285
TD 3 3	**1.7084**	**1.0598**
TD 4 4	.7468	.7378
TD 5 5	**.3664**	**.6473**
TD 6 6	.7515	.8986
TD 7 7	.6512	.6820
TD 8 8	.3945	.4675
TD 9 9	**1.6977**	**.9764**
TD 10 10	.5388	.5618
TD 11 11	**.3972**	**.5930**
TD 12 12	.7956	.7951
TD 13 13	**.3141**	**.5197**
TD 14 14	.4641	.4690
TD 15 15	**.4968**	**.7154**
TD 16 16	.5089	.5192
TD 17 17	**.3079**	**.4718**
TD 18 18	**.3368**	**.5268**
TD 19 19	**1.0189**	**.7285**
TD 20 20	**.2604**	**.3941**
TD 21 21	**.4024**	**.6830**
TD 22 22	**.2967**	**.4857**
TD 23 23	**.2011**	**.3803**

[a] TD is the LISREL syntax for theta-delta

(Those bolded figures are non-invariant items)

Difference in Perception Dispersion. Differences in perception dispersion were identified by constraining all factor variances to be equal between the two groups ($\phi_{1,1}^{(c)} = \phi_{1,1}^{(e)}$ and $\phi_{2,2}^{(c)} = \phi_{2,2}^{(e)}$). This equality constraint again brought a significant $\Delta\chi^2$ ($\Delta\chi^2$ (2) = 30.16, p<.01) plus slight worsening of fit indices. Subsequently only one pair of factor variances was constrained and it was found that employees and customers had different values of factor variances on both factors (table 12). The values of the factor variances were given in table 13. Specifically, it indicated that customers possessed larger values of factor variances on both factors than employees did.

TABLE 12

Results on Testing Differences in Perception Dispersion

Constraints	χ^2	Df	$\Delta\chi^2$	Δdf	p-value	RMSEA	TLI	CFI
PFI	1186.23	472				.065	.922	.927
ALL	1216.39	474	30.16	2	.000*	.065	.919	.925
PH 1 1[a]	1196.21	473	9.98	1	.002*	.065	.921	.927
PH 2 2	1215.94	473	29.71	1	.000*	.066	.919	.925

*p < .01

[a] PH is the LISREL syntax for the phi-matrix

TABLE 13

Values of Factor Variances

Factor Loadings	Employees	Customers
PH 1 1[a]	**.4568**	**.6932**
PH 2 2	**.3817**	**.7001**

[a] PH is the LISREL syntax for the phi-matrix

(Those bolded figures are non-invariant items)

Difference in Baseline Perception. Differences in baseline perception were similarly tested by imposing an equality constraint on all pairs of intercept terms in the CFA model ($\tau_2^{(c)} = \tau_2^{(e)}; \tau_3^{(c)} = \tau_3^{(e)}; \ldots; \tau_{23}^{(c)} = \tau_{23}^{(e)}$). Note that the previously chosen marker items' intercept terms were set to be equal to zero for identification purposes again (Vandenberg & Lance, 2000). This omnibus test brought a significant $\Delta\chi^2$ change ($\Delta\chi^2$ (21) = 180.64, p<.01) plus slight worsening of fit indices. It meant that at least one intercept was significantly different between the two groups. Item-level tests were then executed and the result was shown in table 14. The two groups demonstrated differences in the intercept terms on item 3, 4, 5, 6 (four items of "extrinsic SQ"), 9, 10, 11, 13, 14, 18, 20 and 21 (eight items of "intrinsic SQ").

The values of all the intercept terms were given in table 15. It was found that

out of all twelve non-invariant intercepts, customers had significantly higher intercepts than employees did only on item 3, 9 and 14. They on the other hand had significantly lower intercepts on the rest of the non-invariant items (item 4, 5, 6, 10, 11, 13, 18, 20 and 21). This indicated that customers were often stringent (as manifested in lower starting points of scores) in the SQ evaluations of these "non-invariant" service components than employees.

TABLE 14

Results on Testing Differences in Baseline Perception

Constraints	χ^2	df	$\Delta\chi^2$	Δdf	p-value	RMSEA	TLI	CFI
PFI	1186.23	472				.065	.922	.927
ALL	1366.87	493	180.64	21	.000*	.070	.909	.911
TX 2[a]	1186.70	473	.47	1	.493	.065	.922	.928
TX 3	1195.50	473	9.27	1	.002*	.065	.921	.927
TX 4	1197.13	473	10.90	1	.001*	.065	.921	.926
TX 5	1196.81	473	10.58	1	.001*	.065	.921	.926
TX 6	1195.73	473	9.50	1	.002*	.065	.921	.927
TX 8	1188.95	473	2.72	1	.100	.065	.922	.927
TX 9	1197.85	473	11.62	1	.001*	.065	.921	.926
TX 10	1193.85	473	7.62	1	.006*	.065	.922	.927
TX 11	1198.95	473	12.72	1	.000*	.065	.921	.926
TX 12	1189.53	473	3.30	1	.069	.065	.922	.927
TX 13	1205.95	473	19.72	1	.000*	.065	.920	.926
TX 14	1200.54	473	14.31	1	.000*	.065	.920	.926
TX 15	1186.34	473	.11	1	.740	.065	.922	.928
TX 16	1188.34	473	2.11	1	.146	.065	.922	.927
TX 17	1189.50	473	3.27	1	.071	.065	.922	.927
TX 18	1202.74	473	16.51	1	.000*	.065	.921	.926
TX 19	1191.16	473	4.93	1	.026	.065	.922	.927
TX 20	1192.99	473	6.76	1	.009*	.065	.922	.927
TX 21	1195.13	473	8.90	1	.003*	.065	.922	.927
TX 22	1186.28	473	.05	1	.823	.065	.922	.928
TX 23	1186.79	473	.56	1	.454	.065	.922	.927

*$p < .01$

[a] TX is the LISREL syntax for intercept terms

TABLE 15

Values of Intercept terms

Intercept terms	Employees	Customers
TX 1[a]	0	0
TX 2	.0017	-.0648
TX 3	**-.2075**	**.1206**
TX 4	**-.6401**	**-.9677**
TX 5	**3.8108**	**2.3539**
TX 6	**.5202**	**.2367**
TX 7	0	0
TX 8	.0198	.1666
TX 9	**.4827**	**.8590**
TX 10	**2.0597**	**.7931**
TX 11	**1.9049**	**.4102**
TX 12	.9802	-.0169
TX 13	**1.6482**	**-.0759**
TX 14	**.6519**	**.9671**
TX 15	-.0524	-.0877
TX 16	-.0209	.1139
TX 17	.1420	.2999
TX 18	**1.6588**	**.0563**
TX 19	-.1886	.0351
TX 20	**1.1647**	**.2462**
TX 21	**.4490**	**.1589**
TX 22	1.1595	1.1778
TX 23	.2640	.3275

[a] TX is the LISREL syntax for intercept terms

(Those bolded figures are non-invariant items)

SQ Evaluation

***Difference in Interrelationships among Dimensions*.** As mentioned earlier, the existence of differences in interrelationship among dimensions should be tested by constraining all possible factor correlations to be equal between two groups. There was only one pair of factor correlations in this 2-factor model. The correlation between these two factors was .808 for employees and .782 for customers (table 16). The result of the equality test was given in table 17. It showed that the constraint did not bring a significant $\Delta\chi^2$ ($\Delta\chi^2$ (1) = 0.44, p= .507) and changes in fit indices. Therefore, it was inferred that there was no difference in the perceived

interrelationships among SQ dimensions between these two groups.

TABLE 16

Values of Factor Correlations

Factor correlation	Employees	Customers
CORR 1 2	.808	.782

[a] CORR 1 2 is used to represent factor correlation between the first dimension (extrinsic SQ) and the second one (intrinsic SQ)

TABLE 17

Results on Testing Differences in Interrelationships among Dimensions

Constraints	X^2	df	$\Delta\chi^2$	Δdf	p-value	RMSEA	TLI	CFI
PFI	1186.23	472				.065	.922	.927
CORR 1 2	1186.67	473	.44	1	.507	.065	.922	.928

[a] CORR 1 2 is used to represent factor correlation between the first dimension (extrinsic SQ) and the second one (intrinsic SQ)

Difference in Perceived SQ level. Finally, differences in perceived SQ levels were tested by comparing the means of LVS of like factors as mentioned before. As a comparison, testing using summated average score (SAS) was also performed. The result showed that both SQ dimensions (extrinsic and intrinsic SQ) were found to be differently perceived by employees and customers, with employees' scores always higher than that of customers (regardless of which type of composite score was used; table 18).

The effect size of the mean difference (which was calculated by dividing the mean difference by the pooled standard deviation of the two groups) however showed a considerable difference between these two types of composite score (table

18). Recall that the both dimensions contained non-invariant items between employees and customers, with "extrinsic SQ" contained one non-invariant item and "intrinsic SQ" contained six. The LVS, which took into account non-invariant factor loadings, should be able to reveal a "truer" effect-size of mean difference for these two dimensions. For the dimension "extrinsic SQ", using SAS gave an effect size of .706 whereas LVS gave .747. Thus, it showed that SAS had suppressed the effect size of the mean difference. The situation was more obvious for the dimension "intrinsic SQ" which contained more non-invariant items. Using SAS, the effect size of mean difference was .989 while it was 1.268 using LVS. Again, SAS had suppressed the effect size of the mean difference by a considerable degree (table 18).

TABLE 18

Results on Testing Differences in Perceived SQ level

Dimensions	Score	Employees Mean (s.d.)	Customers Mean (s.d.)	Mean Difference	t-value	*p*	Effect Size
Extrinsic SQ	SAS	6.034 (.7202)	5.419 (.8850)	.615	10.809	.000*	.706
	LVS	5.651 (.6758)	5.034 (.8326)	.617	11.546	.000*	.747
Intrinsic SQ	SAS	6.212 (.6130)	5.336 (.8774)	.876	16.563	.000*	.989
	LVS	6.315 (.6178)	5.094 (.8367)	1.221	23.689	.000*	1.268

*p < .01

Summary of Results

The data set helped investigate if any form of discrepancy emerged between employees and customers in the area of *SQ conceptualization*, *psychometric properties* of the measurement scale and *SQ evaluation*. The conceptual meaning and the possible causes of each form of discrepancy, as well as the empirical findings of the current study are summarized in table 19.

TABLE 19

Seven Forms of Discrepancy between Employees and Customers

Forms of discrepancy	Conceptual Meaning	Possible Causes	Current Study's Findings
Difference in conceptual model	Difference in SQ assessment criteria	• Different knowledge of the service • Different personal service philosophies • Different past experience • Different best-brands taken as referents • Discrepancy between SQ delivered and external communication	No difference. Customers and employees concurrently shared a 2-factor model. The 2 dimensions were extrinsic SQ and intrinsic SQ.
Difference in manifestation of constructs	Difference in strength of relationship between items and dimensions	Same as that of the difference in conceptual model	Customers associated greater strength of relationship between 7 items and the underlying dimensions.
Difference in magnitude of random error	Difference in magnitude of random error in SQ evaluation	• Customers' unfamiliarity with the service • Customers' inattention to survey items • Customers' lapses of memories • Customers mix up with other firms	Customers had a larger magnitude of random errors more often than employees did.
Difference in perception dispersion	Difference in degree of consensus about the SQ level	• Employees soaking in a strong organizational culture • Employees' inconsistent performance • Employees' role ambiguity • Effects of employees training • Customers' various backgrounds	Customers had a lower degree of consensus when comparing to employees on both dimensions.

TABLE 19 (continued)

Seven Forms of Discrepancy between Employees and Customers

Forms of discrepancy	Conceptual Meaning	Possible Causes	Current Study's Findings
Difference in baseline perception	Leniency/stringency in SQ evaluation	• Effects of consumer report • Effects of word-of-moth • Effects of advertisements	Employees were more stringent than lenient in evaluating service components.
Difference in interrelationship among dimensions	Difference in perceived magnitude of interrelationships among the SQ dimensions	• Customers'/employees' unwillingness or inability to discriminate dimensions • Different dominant dimensions • Customers' unfamiliarity with the service • Customers' lack of knowledge of the service	Employees and customers perceived no difference in the relationship between extrinsic and intrinsic SQ.
Difference in perceived SQ level	Employees' over- or under-evaluation of SQ	• Employees' self serving bias (social desirability) • Employees' lack of confidence • Lack of feedback given to employees • Employees' pride of working in the firm	Employees over-evaluated the SQ level of both dimensions when comparing to customers.

SQ Conceptualization. It was found that both groups similarly shared a 2-factor SQ conceptual model in their mindsets as the model fitted both groups well and the two alternative models did not (see table 6). They congruently embraced two dimensions in their SQ conceptualization, namely extrinsic SQ and intrinsic SQ (Mels et al., 1997) as well as associated the same pattern of items to these underlying dimensions. At the same time, they also perceived the same relative strength of relationship between the SQ dimensions and the respective items except for seven particular ones (table 8). Specifically, customers were found to perceive significantly greater strength of relationship between all these seven items and the underlying dimensions than employees did (table 9). The content of these seven service components was given in table 20.

TABLE 20

Service Components of Non-invariant Factor Loadings

Dimension	Item	Service Components
Extrinsic SQ	5[a]	Operating hours convenient to customers
Intrinsic SQ	10	Employees appear neat
	11	Employees provide prompt service
	12	Appearance of employees inspires trust
	13	Employees could place customers as the first priority
	18	Employees are always willing to help customers
	20	Employees are polite

[a] Item 5 corresponds to LX(5,1); item 10 corresponds to LX(10,2) and so forth

Managerial efforts could be directed to synchronize these differences. For instance, knowing that customers perceive a larger strength of relationship between the promptness of employees actions (more than employees do) and "intrinsic SQ" (see table 20), managers could communicate this difference to employees and remind them to abandon any sluggishness possibly shown to customers all the time so as to enhance customers' favorable perception on the "intrinsic SQ" of the service. Given

that there was no difference in conceptual model and only partial difference in manifestation of constructs existed, it was inferred that employees and customers shared a rather similar SQ conceptualization.

Psychometric Properties of the Scale. Differences in magnitude of random error, however, existed on more than half of the items (13 out of 23; table 10). For these items, it was always customers to manifest larger magnitude of random error (table 11), bolstering the idea that customers are usually susceptible to show more random errors in their SQ evaluations, perhaps resulting from the fact that they are usually unfamiliar with all the service components mentioned in the items. Putting this result into application, managers might consider widening the exposure of those service components to customers.

Employees and customers also demonstrated differences in perception dispersion (table 12). Not surprisingly, customers showed larger perception dispersion on both SQ dimensions (table 13). As expounded, managers should be aware if this might be owing to the inconsistent performance of employees. Also, it might evidence that the firm in question was servicing customers from a gamut of different backgrounds. In addition, this might on the other hand suggest that employees were soaking in a strong organizational culture or good training program from which they built the consensus. Managers have to infer from their experience and knowledge what the observed difference particularly means.

In addition, employees and customers exhibited different baseline perception on twelve service components mentioned in the items (table 14). This meant that

customers showed leniency as well as stringency in these SQ components' evaluations. Table 21 provided a summary of the content of these service components.

TABLE 21

Service Components of Non-invariant Intercepts

Dimension	Item	Service Components
Extrinsic SQ	3[a]	Convenient locations
	4	Visually appealing facilities
	5	Operating hours convenient to customers
	6	Up-to-date equipment
Intrinsic SQ	9	Employees respond to customers even when they are busy
	10	Employees appear neat
	11	Employees provide prompt service
	13	Employees could place customers as the first priority
	14	Customers feel safe in their transactions with employees
	18	Employees are always willing to help customers
	20	Employees are polite
	21	When customers have problems, employees are sympathetic and reassuring

[a] Item 3 corresponds to LX(3,1); item 9 corresponds to LX(9,2) and so forth

Customers were found to be stringent more often than being lenient on the evaluations of these service components (table 15). These pieces of information were sufficiently specific and valuable to managers, and could again throw light on how to improve the firm's service. For instance, from the above, it is known that customers might own a higher expectation (as they are more stringent) in employees' politeness when comparing to employees. Managers could on one hand communicate this difference to employees so as to raise their awareness, and on the other hand could provide training to improve employees' courtesy in order to fulfill this higher expectation from customers.

The result of the last three tests of discrepancy showed that employees and customers possessed both similarities and differences in the psychometric properties of the measurement scale they used to evaluate the SQ.

SQ Evaluation. It was found that employees and customers showed no differences in the interrelationships among SQ dimensions (table 17). This might be favorable to the service firm as it suggested that managers could infer customers' perceived interrelationships among SQ dimensions from employees information. Investigating employees' perceived interrelationships among SQ dimensions before prioritizing particular dimensions in marketing seems beneficial.

Finally, employees and customers displayed differences in perceived SQ levels on both SQ dimensions. Like it was surmised, employees held a higher perceived SQ level than customers did on both SQ dimensions (table 18), suggesting to managers that employees were over-evaluating the SQ level they had a part to provide. Possible reasons have to be up to managers' judgment. But as discussed, it might be resulting from the lacking of feedback given to employees or self-serving bias. Despite the underlying reasons, remedial actions should be taken to rectify the situation or the firm might risk losing customers in the end. This test showed that the two groups share a very different view regarding the perceived SQ levels on the SQ dimensions.

The last two tests altogether demonstrated that employees and customers had similarities and differences in their SQ evaluations. In particular, whereas both

parties perceive the same magnitude of interrelationship among SQ dimensions, they perceive very different SQ levels on the SQ dimensions.

CHAPTER VII

DISCUSSION AND CONCLUSION

This study suggests that when examining SQ perception discrepancies existing between employees and customers, researchers and practitioners could actually be able to extract much information pertinent to the two parties by looking into different forms of discrepancy. It is proposed that there can be seven forms of discrepancy emerging from three different areas: SQ conceptualization, psychometric properties of the measurement scale the two groups use to evaluate SQ and SQ evaluation.

Managerial Implication

The empirical result of this study shows that employees and customers do demonstrate several forms of discrepancy in the area of SQ conceptualization, psychometric properties of the measurement scale and SQ evaluation. Although the data was collected in the gas station industry, it still hinted to managers some general similarities and differences between the two parties. First, it was found that both employees and customers concurrently shared a 2-factor conceptual model in their mindsets to conceptualize SQ. These two SQ dimensions were extrinsic SQ and intrinsic SQ. Extrinsic SQ concerned with the quality of the physical facilities used in the service delivery process, whereas intrinsic SQ concerned with the quality of the action and attributes of employees in their service provision (Mels et al., 1997). As suggested by the current study, firms which can place emphasis on the extrinsic (e.g., facilities) as well as the intrinsic aspect (e.g., promptness) of their services are

more likely to fulfill customers' expectation and assessment criteria.

Also, this study provides another piece of evidence to researchers and practitioners that SERVQUAL model might not be an appropriate representation of how people conceptualize SQ (e.g., Carman, 1990). The current study tested a 31-item SERVQUAL model and found that the items did not resemble the SERVQUAL's 5-factor structure. Thus, it may not be fruitful to pursue the development of a standard measurement scale applicable to a wide variety of service industries (Babakus & Boller, 1992). Managers who want to use SERVQUAL must be circumspect that the applicability of SERVQUAL's 5-factor structure in the industry of interest might be limited (Bienstock et al., 1997; Carman, 1990; Durvasula et al., 1999). Managers should not blindly accept SERVQUAL's 5-factor structure. They should instead look for the most interpretable factor structure for their industries and services (Mels et al., 1997).

On the other hand, this study supported the belief that Grönroos's (1984) postulation of the two SQ dimensions, functional quality and technical quality, might be a better representation of how people conceptualize SQ. It was found that employees' and customers' SQ conceptual model could be well represented by intrinsic SQ which resembles functional quality and extrinsic SQ which *to some extent* resembles technical quality, though Grönroos's (1984) interpretation of technical quality is broader (Mels et al., 1997). Technical quality concerns with the outcome (the need satisfied) whereas extrinsic SQ concerns with the tangibles used for need satisfaction (Mels et al., 1997, p.184). Managers might consider using the concept of functional and technical quality to guide them to improve SQ.

Next, it is found that customers tend to show a larger magnitude of random error in their SQ evaluation vis-a-vis employees'. This result demonstrates that customers are usually not familiar enough with every part of the service. Increasing the exposure of the components of the service to customers might be one of the solutions to decrease customers' susceptibility to measurement errors in giving SQ evaluation. Also, this study supports the idea that customers tend to have lower degree of consensus (larger perception dispersion) when evaluating SQ, perhaps owing to the inconsistent performance of employees or the fact that customers are usually from a variety of different backgrounds.

Finally, employees are found to have the inclination to over-evaluate the SQ levels on the SQ dimensions they have a part to contribute when comparing to customers. This alerts to all managers that employees as a matter of fact have a tendency to over-evaluate the perceived SQ level. These empirical results, combined with managers' experience and knowledge about the service, might navigate them to provide more superior SQ to customers.

Using Employees Information. This study also draws managers' attention to comparing employees' and customers' perspective for SQ improvement. Sometimes managers might have focused on customers data only while neglected employees information for generating management decisions and policies directing to customers. Managers however should be noted that comparing employees and customers data before these decisions and policies might further guarantee the positive outcomes. First, a match between the two groups' perspective is likely to bring more satisfied

customers (Klose & Finkle, 1995). Also, as shown in the current study, customers showed a larger magnitude of random error in their SQ evaluation more often than employees did. Therefore, using employees data to support any inferences from customers data might be desirable.

Finally, a mismatched perception between the two groups can be consequential. It should be particularly warned that managers should be alert to any SQ conceptualization discrepancies existing between the two groups as it suggests possible discrepancies in SQ assessment criteria. When employees and customers are found to have different SQ conceptualizations, it might be a reflection that the firms are putting resources at the areas and criteria that customers simply do not care about.

For example, from the empirical result of the current study, it could be known that customers perceived the item "employees could place customers as the first priority" having the strongest relationship with the dimension intrinsic SQ whereas employees perceived that the item "employees could respect as well as enjoy the job" was the one. One of the possible causes for such a difference was that the two groups disagreed on the relative importance of these items as indicators of intrinsic SQ. Customers might think that whether employees could place customers' interest as the very first priority is the most important service attribute for intrinsic SQ. Employees on the other hand might think that whether they could respect and enjoy in the job should be the most important service attribute that determines intrinsic SQ. If this difference goes unnoticed, employees on one hand would keep focusing on an aspect that customers might not consider important to intrinsic SQ and on the other hand

omitting the aspect customers consider the most important. It is therefore maintained that collecting and analyzing employees information in addition to customers data could provide more insights for the enhancement of a firm's well-being than only looking into either side.

Training. Finally, this study suggests to managers that appropriate training is essential for improvement of SQ. As seen from the current study, employees and customers do exhibit differences in several forms of discrepancy in the three areas mentioned, and thus training seems necessary. Training in essence fills the gap between managerial knowledge of customers' expectation and the delivery standard specifications for employees, one of the gaps managers should be attentive to (Zeithaml, Parasuraman, & Berry, 1990). A perennial problem that bothers management teams is giving training in areas that are simply inappropriate or unnecessary (McColl-Kennedy & White, 1997). One of the remedies to this is to acknowledge the possible forms of discrepancy existing between employees and customers (as proposed in this study) and execute corresponding training to employees when such discrepancies are found. A more concrete and detailed understanding of discrepancies between the two groups could prevent the training given to employees from turning futile (McColl-Kennedy & White, 1997).

In addition, in order to deliver superior SQ, a customer-oriented culture must be established and made widespread within the organization nowadays (Daniel & Darby, 1997). Customers-oriented culture in brevity is the set of beliefs that put the customers' interest first (Deshpande, Farley & Webster, 1993). This culture is usually reflected in the attitudes and behaviors of customer-oriented employees (Davidow &

Uttal, 1989) who are usually regarded as part of the service in the eyes of customers. However, managers cannot assume that employees can automatically develop customer-oriented attitudes and behaviors (Daniel & Darby, 1997). Again when managers are armed with a more concrete understanding about the several forms of discrepancy that are existing between the two groups, corresponding training and communication could then be transmitted to employees so as to shape their SQ conceptualization, psychometric properties of the scale they use to evaluate SQ plus SQ evaluation closer to that of customers. This hopefully strengthens the employees' abilities to adjust to their services to take account of the circumstances of the customers (Daniel & Darby, 1997). Therefore, the knowledge of the seven forms of discrepancy proposed in this study is instrumental for managers to achieve a customer-oriented culture.

Methodological Merits

As shown, the seven forms of discrepancy are operationally tested by seven corresponding ME/I tests in SEM. Methodologically, this approach to examine discrepancies between two groups provides researchers with a number of merits (Mackenzie, 2001). First of all, it is well known that SEM controls for the variances due to measurement errors in its estimation of parameters. For example, one of the problems with the presence of measurement errors is that the measurement error in an independent variable artificially attenuates the slope of the relationship between the independent variable and the dependent variable (Mackenzie, 2001). Being able to control for the measurement error, SEM allows researchers and practitioners to interpret the estimates as well as relationships among variables more confidently.

Second, this SEM approach renders a flexibility that allows researchers to test different theoretical factor structures. As mentioned, some studies suggested models to represent SQ conceptualization other than SERVQUAL (see the section of "Conceptualizing SQ). When researchers have reasons to believe that employees and customers are using a particular model to conceptualize SQ, they can test the fitness of the proposed model by SEM and carry on testing other forms of discrepancy between employees and customers. As shown in the current study, three models of different factor structure were tested for fitness in order to look for the one that best represented SQ conceptualization. Further, the approach allows researchers and practitioners to model SQ as a higher-order structure (e.g., Brady & Cronin, 2001; Dabholkar et al., 1996).

The third methodological merit of this SEM approach is its ability to let researchers examine the discrepancies between employees and customers in SQ conceptualization, psychometric properties of the measurement scale and SQ evaluation in *one* way. Researchers seldom examine the discrepancies in the mentioned areas in one study. Even if some researchers had administered an established SQ conceptualization scale (SERVQUAL) in their studies, they still did not compare the SQ conceptualization or the psychometric properties of the measurement scale but only compared the mean scores of each pair of same items (e.g., Hopkins et al, 1993; McColl-Kennedy & White 1997), perhaps being baffled by the lacking of methodology to do so. With the help of seven ME/I tests, the seven forms of discrepancy can be neatly identified with one procedure.

In addition, this SEM approach allows comparison of the results of different studies more efficiently and effectively. If two (or more) studies investigating the discrepancies between employees and customers have concurrently used this SEM approach, specific comparisons can be made between the empirical results of these two studies on each form of discrepancy. This is instrumental for researchers and practitioners to collect neat and systematic information about the discrepancies existing between employees and customers.

Finally, as it is shown, this SEM approach allows the estimation of a new feature called latent variable score (LVS) which in essence is better than summated average score (SAS) and latent means (LM) in comparing the perceived SQ level on a dimension. As between-group means comparison is the lifeblood of most sciences, incessant advancement in the methodology in comparing means seems necessary. The usage of LVS hopefully could stimulate such an advancement.

Measurement Non-invariance as a Source of Information

ME/I tests are the core of this paper as all the proposed discrepancies predicate on the "conceptual meaning" of ME/I tests. In the ME/I literature, the issue of the treatment of non-invariant items (the items that two groups have different values of factor loadings, or in general, parameter estimates) is attracting increasing attentions. For example, items of non-invariant factor loadings can be retained in the doctrine of partial factorial invariance (Byrne et al., 1989). In the past, non-invariant items were always be regarded as a harm to valid inferences (e.g., Vandenberg & Lance, 2000) and therefore should be treated cautiously. However, Cheung and Rensvold (2001)

propose that paramount information could actually be obtained from the non-invariant items. Important between-group differences might only appear in the non-invariant items (Cheung & Rensvold, 2001), especially in cross-cultural studies. This study bolsters this notion that non-invariant items can in fact be informative.

As explained, differences in manifestation of constructs are observed when employees and customers perceive a different strength of relationships between the items and the underlying dimension (manifested in different values of factor loadings). The non-invariant item might matter-of-factly be the ultimate interest of mangers, in that they crave to know what are perceived differently between the two groups so that corresponding actions or strategies can be implemented. Similarly, scalar non-invariance can hint to managers on which service components customers have higher or lower expectation than employees possess. Thus, measurement non-invariance can indeed be viewed as informative rather than harmful.

Further, this study advances the ME/I field by applying ME/I tests into daily life problems, i.e., investigating discrepancies between employees and customers in three aforementioned areas, after they have been applied in a range of context. For instance, ME/I tests have been used to document multiple forms of self-others rating disagreement (Cheung, 1999). Also, ME/I tests, in particular the factorial invariance and scalar invariance test can help identify two types of cultural response bias, namely extreme response style (ERS) and acquiescence response style (ARS) (Cheung & Rensvold, 2000); In addition, multiple tests of ME/I have been shown to be able to document the several types of longitudinal changes (Cheung & Ng, 2001). Its ability to be applied in a range of context roots in the fact that every ME/I test has

its own conceptual meaning and therefore can become the variables of interest in research studies.

Future Direction: Application of the Multiple Forms of Discrepancy

Studies comparing perception discrepancy across groups are abundant in the social science. The multiple forms of discrepancy proposed in this study should not be taken as only for the employees-customers interface. Whenever the interest is in comparing the perceptions of two (or more) parties, these seven forms of discrepancy apply too. When two parties perceive the same thing (especially on some latent variables), chances are there might be discrepancies emerging between the two groups from the three areas discussed. The multiple form of discrepancy proposed in this paper can be adopted to see what kinds of differences actually exist. Do they differ in the conceptual model? Or do they differ in the degree of consensus? These questions can be answered with the analysis proposed here. For example, the multiple forms of discrepancy can be deployed to contrast the conceptualization of supervisors and subordinates, or contrast a group's (e.g., newcomers') psychometric properties of the measurement scale between two points of time, or between male and female on a range of perception variables (e.g., organizational commitment) that researchers and practitioners are interested in comparing.

In addition, as interspersed, this method to examine discrepancy is not limited to two-group comparison. The technique might also be applied to compare more than two groups' perceptions. For example, it can be used to identify any discrepancies existing among the customers (or employees) in different cultures or nations. It can

also be used to compare the conceptualization, or psychometric properties of measurement scale, or the evaluation simultaneously across customers, employees and managers, the three interfaces that commonly interest researchers (e.g., Hartline & Ferrell, 1996; Zeithaml, Parasuraman & Berry, 1990).

Conclusion

Provision of satisfying SQ has become a sine qua non nowadays in order to retain customers in the increasingly competitive market. Firms that could fathom customers' and their employees' SQ conceptualization, psychometric properties of measurement scale they use to evaluate SQ and SQ evaluation, as well as achieve a close match between these two perspectives in these three areas are bound to be more successful than those which cannot. The objective of this study is to apply measurement equivalence/invariance (ME/I) tests to document multiple forms of discrepancy existing between employees and customers in these three areas.

Also as a response to the call to use latent variable structural equation modeling (LVSEM) more in consumer research by Mackenzie (2001), this paper provides a SEM approach to examine these multiple forms of discrepancy between employees and customers. This method on one hand provides neat and rich information to managers about the two parties' perceptions and on the other hand capitalizes the benefits of using SEM. The acknowledgement of these multiple forms of discrepancies between employees and customers could help managers to improve their firms' SQ. On the other hand, the ignorance of these multiple forms of possible discrepancies emerging between the two groups might deteriorate a firm's business performance and competence.

APPENDIX

1. Survey Instrument (in Chinese): SERVQUAL (Parasuraman et al., 1988)

A. Employees Survey

甲部：

請根據你所服務的加德士油站回答以下問題。例如你所服務的是加德士公主道加油站，那麼就請以該加油站作爲回答問題的根據。在回答問題時，你只需要把最適合的答案圈出，1 代表你非常不同意某一句說話，而 7 代表非常同意。你的回答只會被用在整體統計上，公司肯定無法得知你的個人身份和意見。

	非常不同意						非常同意
1. 員工不會因過份忙碌而忽略顧客的要求。	1	2	3	4	5	6	7
2. 在大多數情況下，加油站都會優先考慮到顧客的利益和需要。	1	2	3	4	5	6	7
3. 員工經常保持整潔的外表。	1	2	3	4	5	6	7
4. 員工能爲顧客提供快捷的服務。	1	2	3	4	5	6	7
5. 員工的外貌能夠給與顧客服務信心。	1	2	3	4	5	6	7
6. 員工能夠經常保持「顧客第一」的服務精神。	1	2	3	4	5	6	7
7. 顧客可以安心讓加油站的員工爲他們提供服務。	1	2	3	4	5	6	7
8. 對於顧客所提出的問題和意見，員工都能耐心作出清楚的回應。	1	2	3	4	5	6	7
9. 員工具有敬業樂業的精神。	1	2	3	4	5	6	7
10. 員工能和顧客保持良好的關係。	1	2	3	4	5	6	7
11. 加油站的位置十分方便易找。	1	2	3	4	5	6	7
12. 員工隨時隨地都樂意爲顧客提供任何幫助。	1	2	3	4	5	6	7
13. 加油站擁有吸引的外觀。	1	2	3	4	5	6	7
14. 加油站的營業時間屬於合理和方便。	1	2	3	4	5	6	7

	非常不同意						非常同意
15. 加油站能夠滿足個別顧客的特別要求。	1	2	3	4	5	6	7
16. 加油站能夠兼顧到不同類型顧客的需要。	1	2	3	4	5	6	7
17. 員工經常保持對顧客誠懇有禮。	1	2	3	4	5	6	7
18. 當顧客遇上疑難時，員工會表示關心及幫助他們解決問題。	1	2	3	4	5	6	7
19. 加油站擁有現代化的設施。	1	2	3	4	5	6	7
20. 員工能在合理時間內完成每一項服務。	1	2	3	4	5	6	7
21. 顧客通常不用久候便能獲得服務。	1	2	3	4	5	6	7
22. 員工能夠經常保持熱誠的服務態度。	1	2	3	4	5	6	7
23. 員工具有足夠能力去解答顧客的問題。	1	2	3	4	5	6	7
24. 加油站能履行在各方面的服務承諾。	1	2	3	4	5	6	7
25. 員工的行爲舉止能令顧客對他們具有信心。	1	2	3	4	5	6	7
26. 加油站能夠經常保持四周環境清潔舒適。	1	2	3	4	5	6	7
27. 加油站能提供妥當無誤的服務。	1	2	3	4	5	6	7
28. 加油站的設計能夠迎合顧客的要求。	1	2	3	4	5	6	7
29. 對顧客來說，駕車出入加油站十分方便容易。	1	2	3	4	5	6	7
30. 在任何時間，加油站都有足夠的員工來服務顧客。	1	2	3	4	5	6	7
31. 加油站能爲顧客作出清楚的服務指引。	1	2	3	4	5	6	7

乙部：

個人資料

23. 請問你在加德士公司當什麼職位?

代理商☐　　站長☐　　收銀員☐　　加油員☐

24. 年齡:

20 或以下 ☐　21-25 ☐　26-30　☐

31-35 ☐ 36-40 ☐ 41-45 ☐

46-50 ☐ 51-55 ☐ 56 或以上 ☐

25. 性別： 男☐ 女☐

26. 你在加德士已工作了多少年?

少於 1 年 ☐ 少於 2 年又多於 1 年☐

少於 3 年又多於 2 年 ☐ 少於 4 年又多於 3 年☐

少於 5 年又多於 4 年 ☐ 5 年或以上 ☐

B. Customers Survey

甲部：

請根據你所經常惠顧的加德士油站回答以下問題。例如你所惠顧的是加德士公主道加油站，那麼就請以該加油站作爲回答問題的根據。在回答問題時，你只需要把最適合的答案圈出，1 代表你非常不同意某一句說話，而 7 代表非常同意。你的回答只會被用在整體統計分析上，香港加德士有限公司肯定無法得知你的個人身份和意見。

	非常不同意						非常同意
1. 員工不會因過份忙碌而忽略顧客的要求。	1	2	3	4	5	6	7
3. 在大多數情況下，加油站都會優先考慮到顧客的利益和需要。	1	2	3	4	5	6	7
3. 員工經常保持整潔的外表。	1	2	3	4	5	6	7
4. 員工能爲顧客提供快捷的服務。	1	2	3	4	5	6	7
5. 員工的外貌能夠給與顧客服務信心。	1	2	3	4	5	6	7
6. 員工能夠經常保持「顧客第一」的服務精神。	1	2	3	4	5	6	7
7. 顧客可以安心讓加油站的員工爲他們提供服務。	1	2	3	4	5	6	7
8. 對於顧客所提出的問題和意見，員工都能耐心作出清楚的回應。	1	2	3	4	5	6	7
9. 員工具有敬業樂業的精神。	1	2	3	4	5	6	7
10. 員工能和顧客保持良好的關係。	1	2	3	4	5	6	7
11. 加油站的位置十分方便易找。	1	2	3	4	5	6	7
12. 員工隨時隨地都樂意爲顧客提供任何幫助。	1	2	3	4	5	6	7
13. 加油站擁有吸引的外觀。	1	2	3	4	5	6	7
14. 加油站的營業時間屬於合理和方便。	1	2	3	4	5	6	7
15. 加油站能夠滿足個別顧客的特別要求。	1	2	3	4	5	6	7

	非常不同意						非常同意
16. 加油站能夠兼顧到不同類型顧客的需要。	1	2	3	4	5	6	7
17. 員工經常保持對顧客誠懇有禮。	1	2	3	4	5	6	7
18. 當顧客遇上疑難時，員工會表示關心及幫助他們解決問題。	1	2	3	4	5	6	7
19. 加油站擁有現代化的設施。	1	2	3	4	5	6	7
20. 員工能在合理時間內完成每一項服務。	1	2	3	4	5	6	7
21. 顧客通常不用久候便能獲得服務。	1	2	3	4	5	6	7
22. 員工能夠經常保持熱誠的服務態度。	1	2	3	4	5	6	7
23. 員工具有足夠能力去解答顧客的問題。	1	2	3	4	5	6	7
24. 加油站能履行在各方面的服務承諾。	1	2	3	4	5	6	7
25. 員工的行爲舉止能令顧客對他們具有信心。	1	2	3	4	5	6	7
26. 加油站能夠經常保持四周環境清潔舒適。	1	2	3	4	5	6	7
27. 加油站能提供妥當無誤的服務。	1	2	3	4	5	6	7
28. 加油站的設計能夠迎合顧客的要求。	1	2	3	4	5	6	7
29. 對顧客來說，駕車出入加油站十分方便容易。	1	2	3	4	5	6	7
30. 在任何時間，加油站都有足夠的員工來服務顧客。	1	2	3	4	5	6	7
31. 加油站能爲顧客作出清楚的服務指引。	1	2	3	4	5	6	7

乙部：

個人資料

23. 請問你駕駛以下哪種汽車惠顧加德士加油站?

私家車☐　　的士☐　　小巴☐　　輕型貨車☐　　重型貨車☐

24. 年齡:

20 或以下 ☐ 21-25 ☐ 26-30 ☐

31-35 ☐ 36-40 ☐ 41-45 ☐

1-50 ☐ 51-55 ☐ 56 或以上 ☐

25. 性別： 男☐ 女☐

26. 每個月收入?

$10,000 或以下	☐	$10,001- $20,000	☐
$20,001- $30,000	☐	$30,001- $40,000	☐
$40,001- $50,000	☐	$50,001 或以上	☐

APPENDIX 2

Item Patterns of the Three Testing Models

Model	Dimensions	Survey Items
31-item SERVQUAL	Tangibles	3, 5, 11, 13, 19, 26
	Reliability	18, 20, 24, 27, 31
	Responsiveness	1, 4, 8, 12, 21, 30, 22
	Assurance	6, 7, 9, 10, 17, 23, 25
	Empathy	2, 14, 15, 16, 28, 29
31-item Unidimensional	Service Quality	All 31 items
24-item 2-factor model (Mels et al., 1997)	Extrinsic SQ	28, 29, 11, 13, 14, 19, 26
	Intrinsic SQ	23, 25, 1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 17, 18, 20, 22
23-item 2-factor model (Mels et al., 1997)	Extrinsic SQ	Same as the previous model except taking away item 26
	Intrinsic SQ	Same as the previous model

The numbers in the column "Survey Items" correspond to the items appearing on the survey. See Appendix 1 for the survey items.

REFERENCES

Babakus, E., & Boller, G. (1992). An empirical assessment of the SERVQUAL scale. *Journal of Business Research*, 24, 253-268.

Banks, C. G., & Murphy, K. R. (1985). Toward narrowing the research-practice gap in performance appraisal. *Personnel Psychology*, 38, 335-345.

Bentler, P. M. (1990). Comparative fit indexes in structural equation models. *Psychological Bulletin*, 107, 238-246.

Bienstock, C. C., Mentzer, J., & Bird, M. M. (1997). Measuring physical distribution service quality. *Journal of the Academy of Marketing Science*, 25(1), 31-44.

Bitner, M. J., Booms, B. H., & Mohr, L. A. (1994). Critical service encounters: The employee's viewpoint. *Journal of Marketing*, 58, 95-106.

Bitner, M. J., & Hubbert, A. (1994). Encounter satisfaction versus overall satisfaction versus quality: The customer's voice. In R. Rust, & R. Oliver (Eds.), *Service quality: New directions in theory and practice*. London: Sage, 79-94.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bolton, R. N., & Drew, J. H. (1991). A multistage model of customers' assessment of service quality and value. *Journal of Consumer Research*, 17, 375-384.

Brady, M. K., & Cronin, J. J. (2001). Some new thoughts on conceptualizing perceived service quality: A hierarchical approach. *Journal of Marketing*, 65, 34-49.

Brensinger, R. P., & Lambert, D. M. (1990). Can the SERVQUAL scale be generalized to business-to-business services? In *Knowledge Development in Marketing*, AMA's Summer Educators Conference Proceedings, Boston, MA, 1990, p. 289.

Brown, T. J., Churchill Jr., G. A., & Peter, J. P. (1993). Improving the measurement of service quality. *Journal of Retailing*, 69(1), 127-139.

Brown, S.W., & Swartz, T. A. (1989). A gap analysis of professional service quality. *Journal of Marketing*, 53, 92-98.

Bruhn, M., & Georgi, D. (2000). Information-based analysis of service quality gaps. Managing service quality by internal marketing. *Journal of Professional Services Marketing*, 21(2), 105-124.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.

Cadotte, E. R., Woodruff, R. B., & Jenkins, R. L. (1987). Expectations and norms in models of consumer satisfaction. *Journal of Marketing Research*, 24, 305-314.

Carman, J. M. (1990). Consumer perceptions of service quality: An assessment of the SERVQUAL dimensions. *Journal of Retailing*, 66, 33-55.

Caruana, A., Weing, M. T., & Ramaseshan, B. (2000). Assessment of the three-column format SERVQUAL: An experimental approach. *Journal of Business Research*, 49(1), 57-65.

Chan, D. (1997). Racial subgroup differences in predictive validity perceptions on personality and cognitive ability tests. *Journal of Applied Psychology*, 82(2), 311-321.

Cheung, G. W. (1999). Multifaceted conceptions of self-other ratings disagreement. *Personnel Psychology*, 52, 1-36.

Cheung, G. W., & Ng, T. W. (2001). *Multifaceted conceptions of change: Revisiting the alpha, beta and gamma change typology*. Paper presented at the meeting of Academy of Management, Washington D. C.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, 25(1), 1-27.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence

response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31(2), 187-212.

Cheung, G. W., & Rensvold, R. B. (2001). *Comparison of latent means with structural equation modeling: A latent variable approach.* Paper presented at the meeting of Academy of Management, Washington D. C.

Comila, S., (2000). Employee and customer perceptions of service quality: A look at India. *Current Psychology*, 19(4), 292-301.

Cronin, J., & Taylor, S. A. (1992a). Measuring service quality: A re-examination and extension. *Journal of Marketing*, 58(1), 125-131.

Cronin, J., & Taylor, S. A. (1992b). SERVPERF versus SERVQUAL: Reconciling performance based and perceptions-minus-expectations measurement of service quality. *Journal of Marketing,* 58(1), 125-131.

Daniel, K., & Darby, D. N. (1997). A dual perspective of customer orientation: A modification, extension and application of the SOCO scale. *International Journal of Service Industry Management*, 8(2), 131-147.

Dabholkar, P. A., Shepherd, C. D., & Thorpe, D. I. (2000). A comprehensive framework for service quality. An investigation of critical conceptual and measurement issues through a longitudinal study. *Journal of Retailing*, 76(2), 139-173.

Dabholkar, P. A., Thorpe, D. I., & Rentz, J. O. (1996). A measure of service quality for retail stores. *Journal of the Academy of Marketing Science*, 24, 3-16.

Davidow, W. H., & Uttal, B. (1989). *Total customer service: The ultimate weapon.* Haper Collins, New York, NY.

Deshpande, R., Farley, J. U., & Webster, F. E. Jr (1993). Corporate culture, customer orientation, and innovativeness in Japanese firms: A quadrad analysis. *Journal of Marketing*, 57, 23-27.

Durvasula, S., Lysonski, S., & Mehta, S. C. (1999). Testing the SERVQUAL scale in the business-to-business sector: The case of ocean freight shipping service.

Journal of Service Marketing, 13(2), 132-150.

Finn, D. W., & Lamb, C. W. (1991). An evaluation of the SERVQUAL scales in a retailing setting. *Advances in Consumer Research*, 18, 338-357.

Fok, L. Y., Hartman, S. J., Villere, M. F., & Freibert, R. C. III (1996). A study of the impact of cross cultural differences on perceptions of equity and organizational citizenship behavior. *International Journal of Management*, 13(1), 3-15.

Franke, G. R., Crown, D. F., & Spake, D. F. (1997). Gender differences in ethical perceptions of business practices: A social role theory perspective. *Journal of Applied Psychology*, 82(6), 920-934.

Gagliano, K. B., & Hathcote, J. (1994). Customer expectations and perceptions of service quality in retail apparel specialty stores. *Journal of Services Marketing*, 8(1), 60-69.

Genestre, A., & Herbig, P. (1996). Service expectations and perceptions revisited: Adding product quality to SERVQUAL. *Journal of Marketing Theory and Practice*, 4(4), 72-82.

Grapentine, T. (1998). The history and future of service quality assessment. *Marketing Research*, 10(4), 4-17.

Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of Marketing*, 18(4), 36-44.

Hartline, M. D., & Ferrell, O. C. (1996). The management of customer-contact service employees: An empirical investigation. *Journal of Marketing*, 60, 52-70.

Hansen, D. E., & Danaher, P. J. (1999). Inconsistent performance during the service encounter. *Journal of Service Research*, 3, 227-235.

Hopkins, S. A., Strasser, S., Hopkins, W. E., & Foster, J. R. (1993). Service quality gaps in the transportation industry. *Journal of Business Logistics*, 14(1), 145-162.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to

measurement invariance in aging research. *Experimental Aging Research*, 18, 117-144.

Iacobucci, D., Grayson, K. A., & Ostrom, A. L. (1994). The calculus of service quality and customer satisfaction: Theoretical and empirical differentiation and integration. In T. A. Swartz, D. A. Bowen, & S. W. Brown (Eds), *Advances in Services and Marketing and Management*, Vol. 3, JAI Press, Greenwich, CT and New York, NY, 1-67.

Iacabucci, D., Ostrom, A., & Grayson, K. (1995). Distinguishing service quality and customer satisfaction: The voice of the consumer. *Journal of Consumer Psychology*, 4(3), 277-303.

Johnson, J. J. (2000). Differences in supervisor and non-supervisor perceptions of quality culture and organizational climate. *Public Personnel Management*, 29(1), 119-129.

Johnston, R. (1995). The determinants of service quality: Satisfiers and dissatisfers. *International Journal of Service Industry Management*, 6(5), 53-71.

Jöreskog, K. (2000). *Latent variable scores and their uses*. Web-site of Scientific Software International.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of Behavioral Research* (Eds.). Harcourt Brace.

Kim, J., & Mueller, C. W. (1994). Factor analysis: Statistical methods and practical issues. In M. S. Lewis-Beck (Ed.) *Factor analysis and related techniques*. CA: Sage.

Klose, A., & Finkle, T. (1995). Service quality and the congruency of employee perceptions and customer expectations: The case of an electrical utility. *Psychology & Marketing*, 12(7), 637-646.

Lam, S. (1997). SERVQUAL: A tool for measuring patients' opinions of hospital service quality in Hong Kong. *Total Quality Management*, 8(4), 145-152.

Lee, H., Lee, Y., & Yoo, D. (2000). The determinants of perceived service quality

and its relationship with satisfaction. *Journal of Services Marketing*, 14(3), 217-231.

Lehtinen, U., & Lehiten, J. R. (1985). *Service quality: A study of quality dimensions*. Second World Marketing Congress, University of Stirling, Scotland.

Leuthesser, L., Kohli, C. S., & Harich, K. R. (1995). Brand equity: The halo effect measure. *European Journal of Marketing*, 26(4), 57-66.

Llosa, S., Chandon, J., & Orsingher, C. (1998). An empirical study of SERVQUAL's dimensionality. *The Service Industries Journal*, 18(2), 16-44.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental tests scores*. Reading, MA: Addison-Wesley.

Mackenzie, S. B. (2001). Opportunities for improving consumer research through latent variable structural equation modeling. *Journal of Consumer Research*, 28, 159-166.

Mattsson, J. (1994). Improving service quality in person-to-person encounters: Integrating findings from a multi-disciplinary review. *The Service Industries Journal*, 14(1), 46-61.

McAlexander, J. H., Kaldenbugy, D. O., & Koenig, H. F. (1994). Service quality measurement. *Journal of Health Care Marketing*, 14(3), 34-42.

McColl-Kennedy, J. R., & White, T. (1997). Service provider training programs at odds with customer requirements in five-star hotels. *The Journal of Service Marketing*, 11(4), 249-264.

McDougall, G. H. G., & Levesque, T. J. (1994). A revised view of service quality dimensions: An empirical investigation. *Journal of Professional Services Marketing*, 11(1), 189-209.

Mels, G., Boshoff, C., & Nel, D. (1997). The dimensions of service quality: The original European perspective revisited. *Service Industries Journal*, 17(1), 173-189.

Mersha, T., & Adlakha, V. (1992). Attributes of service quality: The consumer perspective. *International Journal of Service Industry Management*, 3(3), 34-45.

Morris, T., & Pavett, C. M. (1992). Management style and productivity in two cultures. *Journal of International Business Studies*, 23(1), 169-179.

Mullen, M. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies*, 26, 573-596.

Murphy, K. R. & Jako, R. A. (1989). Under what conditions are observed intercorrelations greater or smaller than true intercorrelations? *Journal of Applied Psychology*, 74(5), 823-830.

Nishida, H., Hammer, M. R., & Wiseman, R. L. (1998). Cognitive differences between Japanese and Americans in their perceptions of difficult social situations. *Journal of Cross-Cultural Psychology*, 29(4), 449-524.

Oliver, R. L. (1993). A conceptual model of service quality and service satisfaction: Compatible goals, different concepts. In T. A. Swartz, D. E. Bowen, & S. W. Brown (Eds.), *Advances in Marketing and Management*. Greenwich, CT: JAI Press, 65-85.

Parasuraman, A., & Berry, L. L. (1991). *Marketing Services: Competing through quality*, Maxwell Macmillan, Canada.

Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1991). Refinement and reassessment of the SERVQUAL scale. *Journal of Retailing*, 67, 420-450.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49, 41-50.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40.

Parasuram, A., Zeithamal, V. A., & Berry, L. L. (1994). Alternative scales for measuring service quality: A comparative assessment based on psychometric

and diagnostic criteria. *Journal of Retailing*, 70, 201-230.

Peter, J., Churchill, G., & Brown, T. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research*, 19(4), 655-662.

Philip, G., & Hazlett, S. (1997). The measurement of service quality: A new P-C-P attributes model. *The International Journal of Quality and Reliability Management*, 14(3), 260-279.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552-566.

Reynierse, J. H. & Harker, J. B. (1992). Employee and customer perceptions of service in banks: Teller and customer service representative ratings. *Human Resource Planning*, 15(4), 31-46.

Robinson, S. (1999). Measuring service quality: Current thinking and future requirements. *Marketing Intelligence and Planning*, 17(1), 21-34.

Ronan, W. W., & Latham, G. P. (1974). The reliability and validity of the critical incident technique: A closer look. *Studies in Personnel Psychology*, 6, 53-64.

Rosan, L. D., & Karwan, K. R. (1994). Prioritizing the dimensions of service quality. An empirical investigation and strategic assessment. *International Journal of Service Industry Management*, 5(4), 39-52.

Rust, R. T., & Oliver, R. L. (1994). Service quality: Insights and managerial implications from the frontier. In R. T. Rust, & R. L. Olivers (Eds.), *Service quality: New directions in theory and practice*. Thousand Oaks, CA: Sage Publications, 1-19.

Schank, R. C., & Abelson, R. P., (1977). Scripts, plans, goals and understanding. New York: John Wiley and Sons Inc.

Schneider, B., & Bowen, D. E. (1985). Employee and customer perceptions of service in banks: Replication and extension. *Journal of Applied Psychology*, 70(3), 423-433.

Schneider, B., Parkington, J. J., & Buxton, V. M. (1980). Employee and customer perceptions of service in banks. *Administrative Science Quarterly*, 25, 252-267.

Schneider, B., Wheeler, J. K., Cox, J. F. (1992). A passion for service: Using content analysis to explicate service climate themes. *Journal of Applied Psychology*, 77(5), 705-716.

Schneider, B., White, S. S., & Paul, M. C. (1998). Linking service climate and customer perceptions of service quality: Test of a causal model, *Journal of Applied Psychology*, 83(2), 150-163.

Soloman, M. R., Suprenant, C., Czepiel, J. A., & Gutman, E. G. (1985). A role theory perspective on dyadic interactions: The service encounter. *Journal of Marketing*, 49, 99-111.

Spreng, R. A., & Mackoy, R. D. (1996). An empirical examination of a model of perceived service quality and satisfaction. *Journal of Applied Psychology*, 72(2), 201-214.

Spreng, R. A., & Singh, A. K. (1993). An empirical assessment of the SERVQUAL scale and the relationship between service quality and satisfaction. In *Knowledge Development in Marketing*, AMA's Summer Educators Conference Proceedings, Boston, MA, 1993, p. 1-6.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.

Teas, R. K. (1993). Expectations , performance evaluation, and consumers' perception of quality, *Journal of Marketing*, 57(4), 18-34.

Tucker, L. R., & Lewis, C. A. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the

measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-69.

Vandenberg, R. J., & Self, R. M. (1993). Assessing newcomers' changing commitment to the organization during the first 6 months of work. *Journal of Applied Psychology*, 78, 557-568.

Wirtz, J., & Bateson, J. E. G. (1995). An experimental investigation of halo effects in satisfaction measures of service attributes. *International Journal of Service Industry Management*, 6(3), 84-102.

Yavas, U. (1998). Further evidence on the psychometric properties of SERVQUAL in Turkey: A replication and extension. *Journal of International Marketing and Marketing Research*, 23, 59-70.

Zeithaml, V. A. (1987). Defining and relating price, perceived quality, and perceived value, Report No. 87-101, Cambridge, MA: *The Strategic Planning Institute*, PIMSLETTER No. 33.

Zeithaml, V. A., (2000). Service quality, profitability, and the economic worth of customers: What we know and what we need to learn. *Journal of Academic of Marketing Science*, 28(1), 67-85.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996) The behavioral consequences of service quality. *Journal of Marketing*, 60, 31-46.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1988). Communication and control processes in the delivery of service quality. *Journal of Marketing*, 52, 35-48.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1993). The nature and determinants of customer expectations of service. *Journal of the Academy of Marketing Science*, 21(1), 1-12.

Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1990). *Delivering quality service*. New York : Free Press ; London : Collier Macmillan.